Are Conditional Cash Transfers Fulfilling Their Promise? Schooling, Learning, and Earnings After 10 Years

By TANIA BARHAM, KAREN MACOURS, AND JOHN A. MALUCCIO[1]

September 2018

Abstract: Interventions promoting investment in child human capital are motivated by their potential to break the intergenerational transmission of poverty. With this goal, conditional cash transfer (CCT) programs are the anti-poverty program of choice in many developing countries even though the evidence on their longer term objectives is inconclusive. This paper uses the randomized phase-in of a CCT program in Nicaragua to estimate long-term education, learning, and labor market effects for males 10 years after the start of the program. Gains in schooling and learning coincide with positive labor market returns including higher earnings, and demonstrate important long-term returns to CCTs.

JEL Codes: I25, I38, I28

Key words: CCT, long-term effects, education, learning, labor markets

---

[1] Barham: Department of Economics and IBS, University of Colorado at Boulder, Boulder, CO 80309-0256 (tania.barham@colorado.edu); Macours: Paris School of Economics and INRA, 48 Boulevard Jourdan, 75014 Paris, France (karen.macours@psemail.eu); Maluccio, Department of Economics, Middlebury College, Middlebury VT 05753 (maluccio@middlebury.edu). Acknowledgments: This research would not have been possible without the support of Ferdinando Regalia of the Inter-American Development Bank (IDB). We gratefully acknowledge generous financial support from IDB, the Initiative for International Impact Evaluation (3ie: OW2.216), and the National Science Foundation (SES 11239945 and 1123993). See also AEA RCT registry AEARCTR-0001572. We are indebted to Veronica Aguilera, Enoe Moncada, and the survey team from CIERUNIC for excellent data collection and for their persistence in tracking. We also acknowledge members of the *Red de Protección Social* program team (in particular, Leslie Castro, Carold Herrera, and Mireille Vijil) for discussions regarding this research and Emma Sanchez Monin for facilitating the data collection process. We thank Teresa Molina Millán, Olga Larios, Jana Parsons, and Gisella Kagy for help with data preparation and Vincenzo di Maro for numerous contributions to the tracking survey. Finally, we are grateful for comments received from Norbert Schady, Guillermo Cruces and others during presentations at the IDB, Northeast Universities Development Conference 2012, Allied Social Sciences Association 2013, Colby College, Middlebury College, European Development Network Conference 2013, Population Association of America meetings, UCL, Louvain-la-Neuve, Pompeu Fabra, and at the Impact Evaluation Network conference of LACEA. This paper modifies and extends the IADB working paper 432 "More Schooling and More Learning?: Effects of a 3-Year Conditional Cash Transfer Program in Nicaragua After 10 Years." All remaining errors and omissions are our own.

Interventions aimed at increasing the nutrition, health, and education of children are often motivated by their potential to break the intergenerational transmission of poverty. Rigorous evidence regarding the short-term effects of such interventions has expanded tremendously, especially in low- and middle-income countries. Relatively little, however, is known regarding whether such interventions live up to their full promise, and increase the productive potential of the next generation.[2] The lack of a solid evidence base is due in part to the methodological challenges facing long-term assessment and to evaluations designed to measure short-term effects that are often ill suited for establishing long-term gains, or that may still be too recent.

Consideration of long-term impacts is particularly relevant for conditional cash transfer (CCT) programs. Started in 1997 in Mexico and Brazil, CCTs have spread to more than 60 countries and reach almost one-quarter of the population, or 135 million individuals, in Latin America alone (World Bank 2015; Robles, Rubio, and Stampini 2018). The principal program components of the prototypical CCT include regular cash transfers to women—conditional on scheduled visits to healthcare providers for young children and on school enrollment and regular attendance for school-age children—and social marketing to encourage investment in nutrition, health, and education. Numerous evaluations, many based on experimental designs, have shown positive short-term effects. These include poverty alleviation, improved nutrition and health (particularly for young children), and increased school attainment (Fiszbein and Schady 2009). Short-term schooling effects have largely been sustained in the long-term, but evidence of the longer-term effects on other outcomes is mixed and inconclusive (Molina Millán et al. 2018a).

We contribute to filling this evidence gap by analyzing the long-term effects of a randomized CCT program in Nicaragua on human capital and labor market outcomes for boys, in a setting in which we can comprehensively address a number of methodological challenges inherent to research on long-term impacts. We demonstrate that exposure to the CCT at critical ages (when boys are likely to drop out of primary school) not only led to sustained educational and learning gains of nearly 0.2 standard deviations but also to increased returns in the labor market. The young men were more likely to engage in wage work, migrate temporarily for better paying jobs, and earn more, with nearly 15 percent higher earnings per month worked. Cost-benefit analysis suggests the program could achieve a positive net present value within two decades.

[2] Important exceptions include evidence on the long-term effects of early childhood stimulation (Gertler et al. 2014), early childhood nutrition (Hoddinott et al. 2008), deworming (Baird et al. 2016), education subsidies and HIV prevention education (Duflo, Dupas, and Kremer 2015) and school vouchers (Bettinger et al. 2016).

The main analysis exploits the randomized phase-in of the program and its eligibility rules to estimate intent-to-treat (ITT) effects after 10 years. Households randomly assigned to the early treatment group were eligible for transfers from 2000 to 2003, while those assigned to late treatment were eligible from 2003 to 2005. In both groups, all households were eligible for transfers for nutrition and health, but only households with children between 7 and 13 were eligible for education transfers. Maluccio and Flores (2005) established that the program led to relatively large short-term increases in enrollment and educational attainment. We identify a cohort of boys aged 9-12 at the start of the program in 2000 for whom the randomized phase-in implied they had greater exposure to the education transfers at ages critical for schooling in the early treatment group. Specifically, due to the timing of exposure, eligibility rules, and pre-program school dropout patterns, early treatment boys were exposed to the education transfers of the program at ages when they were more likely to prevent school dropout than boys exposed in the late treatment group. Since the late treatment group became eligible for the program approximately three years after the early treatment group, there is no long-term experimental control group and we cannot estimate corresponding long-term absolute program effects. Instead, we estimate differential ITT effects. While we focus on school age children at program start, the estimated effects reflect all components of the CCT program, not just the education component, as is true for most CCT evaluations.

To establish the differential long-term program effects, we collected data in 2010 on households and individuals originally interviewed prior to randomization and program start in 2000, including an oversample of individuals for whom the exposure differential to the education transfers was large. The oversample helps address concerns regarding limited statistical power in the differential treatment design. Special effort was made to minimize attrition, given the potential (but a priori unknown) relationship between program exposure, education, and migration in this highly mobile age group of young men. Final attrition is relatively low (10 percent for the main outcomes on education, labor market, and earnings, and ranging between 4 and 19 percent for other outcomes) and is balanced between early and late treatment groups, and the resulting analysis samples are balanced on baseline observables. As those who migrated and could not be found may still be different from those who were found, however, we consider a variety of complementary estimation strategies to gauge the importance of any remaining selection. Throughout the analysis we account for multiple hypothesis testing by grouping

outcomes in families, and test the familywise error rate following Anderson (2008). We further use randomization statistical inference to calculate exact p-values for sharp null hypotheses, as recommended by Athey and Imbens (2017) and Young (2017).

We timed the follow-up survey so that learning and labor market outcomes were measured approximately ten years after the start of the program, and seven years after households eligible for the program in the early treatment group stopped receiving transfers. By that time, when the young men were 19-22 years old, virtually all were in the labor force, and most had completed their schooling. Therefore, the analysis offers a rare opportunity to examine the sustainability of CCT program effects, as we assess whether individuals who might have received more schooling because of the program perform better on achievement tests and in their initial years in the labor market, well after the program ended.

The experimental design described above yields the long-term differential program effects. These estimates, while requiring relatively few assumptions for identification, are likely to underestimate the long-term absolute effects of the CCT. This is because boys in the 9-12 cohort in the late treatment group lived in households that were eligible for the program starting in 2003, and hence still may have benefitted from the overall program. Therefore, we complement the experimental differential estimates with non-experimental estimates of the absolute effects using two different approaches and relying on primary data from a separate non-experimental comparison group, as well as data from the Nicaraguan national census. Results are broadly consistent with the experimental estimates and confirm positive program impacts.

While the evidence supports it, the interpretation that the differential effects represent lower bounds still relies on the assumption that the program affected outcomes in both treatment groups in the same direction (though with smaller positive impacts in the late treatment group). This assumption is plausible for boys in this age cohort, for whom long-term labor market impacts most likely result primarily from the impact of the CCT on education. Interpretation of differential estimates, however, is less straightforward for girls in the same cohort, because the nutrition and reproductive health components of the intervention can directly affect the age of menarche and early fertility, which independently from education can affect their labor market outcomes. Given these differences in the likely causal pathways and the interpretation of the differential results, we analyze the girls separately in a companion paper (Barham, Macours, and Maluccio 2018). Importantly, for girls we also find positive and significant differentials on

earnings and labor market outcomes, similar in size to those for the boys. Education differentials, however, are more modest, and there are no differential learning effects. The limited education differentials are consistent with girls dropping out of school later in this context and related educational catch-up by girls in the late treatment group. Instead, for girls there are strong differential effects on fertility outcomes.

The CCT program for which there is the deepest evidence base, in part because it was the first to have an experimental design, is the Mexican program, *PROGRESA*. As such, our analysis closely relates to research on that program, in particular the work of Behrman, Parker, and Todd (2009a, 2011). They exploit an 18-month experimental difference in exposure to an ongoing CCT, complemented by non-experimental matching estimates of absolute program effects after 5.5 years, for youth ages 15–21 at follow-up. Their results indicate that the program increased schooling but not learning. Moreover, they find a small reduction in labor force participation for the young men, a reflection that the age cohort in the analysis is young and may not yet fully show the potential effects of CCTs on labor market outcomes. Parker and Vogl (2018) also examine effects of the Mexican CCT on income using the non-random national rollout and census data 13 years after the start of the program, exploiting a larger difference in exposure and measuring outcomes when individuals were 20-24 years old. For men, they find a shift away from agricultural work and an increase in the probability of working in the formal sector, though no significant increase in earnings, while results for women indicate substantial increases in both labor market participation and earnings.

Beyond Mexico, while positive long-term effects of CCTs on schooling have been found in a variety of other settings with different research designs, evidence on learning and on labor market returns is more limited (Molina Millán et al. 2018a). This is in part because the randomized phase-in of some of the early programs implies the statistical power for long-term evaluations is often be limited. We circumvent this challenge by focusing on a cohort with a relatively large differential exposure to the program at ages critical for school dropout, adding an oversample of individuals in this cohort in the follow-up survey, and examining a program in which early and late treatment households received the program only for a fixed time period, allowing measurement of differential effects after the intervention ended.

This paper closely relates to two experimental studies of CCTs that investigate absolute experimental impacts after more than 10 years in Colombia and Honduras (Barrera-Osorio,

Linden, and Saavedra 2018; Molina Millán et al. 2018b). These studies rely on administrative or census data, show clear long-term schooling gains, but lack sufficient information to fully examine impacts on learning or earnings. Therefore, an important contribution of the current study is to provide evidence for a rich set of outcomes that trace the possible causal pathway from transfers to schooling, learning, labor market activity and earnings. As such, this paper also relates to Baird, McIntosh, and Özler (2016) and Macours and Vakis (2016) who use survey data and experimental variation to estimate absolute program impacts, although only two years after the end of the respective CCT programs.

More generally, we complement evidence from a number of non-experimental studies examining effects over long periods before phase-in of a comparison group (García et al. 2012; Parker and Vogl 2018), but requiring stronger identification assumptions. Assumptions are weaker for studies using regression discontinuity designs. These estimate local average treatment effects (LATE) for the least poor beneficiaries, however, which may differ from overall average treatment effects (Filmer and Schady 2014; Araujo, Bosch, and Schady 2016). Our results indicate that at least in Nicaragua, effects were much larger for the poorest, suggesting studies estimating LATE can miss impacts on groups that benefited the most.

An important source of concern in many studies is sample attrition, often directly related to migration and therefore likely related to labor market outcomes. Studies using secondary administrative data can suffer from related selection concerns when multiple data sources are combined but can only be imperfectly matched (Baez and Camacho 2011). Studies using census data can sidestep these concerns if information on municipality of birth is available, but even those typically still exclude international migrants. To limit selection concerns in our study, migrants were intensively tracked including to Costa Rica, the main international migration destination, so that attrition for important outcomes is lower than in related studies.[3] It turns out that migration patterns are crucial for understanding the labor market returns.

Finally, most longer-term studies of CCTs have focused on beneficiaries while they are still transitioning into the labor market, complicating interpretation of labor market impacts due to the potential tradeoff between additional schooling and shorter work experience. While we also

[3] Attrition rates for our main labor market and earnings outcomes are 10 percent, lower than in related studies on long-term labor market effects of CCTs (see appendix F).

investigate a cohort that is still relatively young, 98 percent of them are already economically active, partly circumventing that concern.

More broadly, this paper complements research on long-term effects of CCT programs beyond learning and labor market outcomes. Gertler, Martinez, and Rubio-Codina (2012) show effects on household investment in economic activities. Several studies analyze longer-term effects of exposure at younger ages, i.e., during early childhood.[4] In addition, this paper relates to recent studies analyzing long-term effects of randomized anti-poverty interventions such as the ultra-poor transfer programs in Bangladesh (Bandiera et al. 2017) and India (Banerjee et al. 2016) or unconditional cash transfers in Kenya (Haushofer and Shapiro 2018).[5]

Finally, and in contrast to much of the existing evidence, our results on learning suggest that at least in some contexts CCTs may have a role to play not only in promoting schooling but also in helping children learn. Filmer and Schady (2014) find no short-term effects of CCTs on tests of mathematics and language in Cambodia. Baird, McIntosh, and Özler (2011) find significant short-term effects on learning in Malawi but no effect in the longer-term (Baird, McIntosh, and Özler 2016). Duque, Rosales-Rueda and Sanchez (2018), however, do find positive long-term learning results in Colombia.

Taking the learning and earning results together, this paper has important policy implications. Indeed, the lack of conclusive evidence on the long-term effects of CCTs has led some to conclude that the potential for CCTs to reduce the intergenerational transmission of poverty may be limited (Levy and Schady 2013). The evidence we present, for a program closely resembling many other CCT programs arguably means there is room for more optimism.

II. The Nicaraguan CCT Program and Its Experimental Design

A. Key program design features of the CCT

Modeled after *PROGRESA*, the *Red de Protección Social* (*RPS*) was a CCT designed to address both current and future poverty through cash transfers targeted to poor households in rural Nicaragua. The government of Nicaragua implemented the program with technical

[4] See Behrman, Parker, and Todd (2009b); Fernald, Gertler, and Neufeld (2009); Macours, Schady, and Vakis (2012); Barham, Macours, and Maluccio (2013); and Araujo, Bosch, and Schady (2016).
[5] See also related research on long-term effects of childhood cash transfers or other income shocks on outcomes in the U.S. and Norway, mostly based on non-experimental variation (Akee et al. 2010; Aizer et al. 2016; Cesarini et al. 2016; Hoynes, Schanzenbach, and Almond 2016; Loken, Moggstad and Wiswall 2012).

assistance and financial support from the Inter-American Development Bank (IDB). On average, transfers were 18 percent of pre-program expenditures and delivered bimonthly (every two months). They were paid to a designated household representative (typically a female caregiver) in the beneficiary household and came with a strong social marketing message that the money was intended to be used for educational, food, and health investments.

The CCT had two core components. The first component focuses on education. Households with children aged 7–13 years who had not yet completed the fourth grade of primary school were eligible for the education components of the program. They received an additional fixed bimonthly cash transfer known as the school attendance transfer, which was contingent on enrollment and regular school attendance of those children. For each eligible child, the household also received an annual cash transfer at the start of the school year, intended for school supplies, conditional on enrollment. We refer to the combined schooling attendance and school supplies transfers as the education transfer. Teachers were required to report enrollment and attendance using forms specifically designed by the program for the verification of the conditions and received a small per-student cash transfer, known as the "teacher transfer".

The second component aimed at improving nutrition and health, and all households were eligible for a transfer of fixed amount per household regardless of the household's size and composition. This nutrition and health transfer was conditional on preventive healthcare visits for any children under age five in the households, and on the household representative attending bimonthly health information workshops. More detailed information on the program is provided in appendix C and Maluccio and Flores (2005).

While the Nicaraguan CCT was modeled after Mexico's original CCT, two differences between the programs are important for our analyses. First, in Nicaragua beneficiary households were only eligible to receive the program for a fixed period of three years, after which it was not possible to renew eligibility. Second, the education conditionalities and transfers in the Nicaraguan CCT only applied to the first four grades of primary school, reflecting the low levels of education and high primary school dropout rates in these impoverished rural areas.

B. Experimental Design of the CCT

The randomized evaluation was built into the design of the CCT intervention in six rural municipalities from three regions in central and northern Nicaragua, chosen based on their low

education and health indicators. In these six municipalities, 42 out of 59 rural *comarcas* (hereafter, localities) were selected based on a marginality index.[6] A program census of all residents in these 42 localities was collected in May 2000.

The targeted localities were then randomized into one of two equally sized treatment groups, the early or late group, at a public lottery. To improve the likelihood that the selection of localities in the experimental groups would be well balanced in terms of poverty levels, the marginality index was used to classify the 42 localities into seven strata of six localities each. From each stratum three localities were randomly selected as early treatment and three as late treatment. The randomization occurred in July 2000, after the program census, and the 21 early treatment localities received their first transfers in November 2000. All but 3 percent of households in these localities were eligible to receive three years' worth of cash transfers, with the last transfer delivered in late 2003. Households in the late treatment localities were informed that the program would start in their localities later. The 21 late treatment localities were phased-in at the beginning of 2003. They were also eligible to receive three years' worth of cash transfers. Households in the early treatment group did not receive any transfers after 2003, and therefore were not affected by any conditionalities after that date. The small teacher transfer, however, continued. At the end of 2005, all program benefits were discontinued for both groups and the program no longer operated in these municipalities.[7]

Overall, compliance with the experimental design was high. Past analysis on the program shows that the sample was balanced at baseline and that there was little contamination of the late treatment group (Maluccio and Flores 2005). Appendix A shows balance for the main sample used in this paper. At the household level, take-up of at least one of the components of the program in early and late treatment localities was approximately 85 percent.

III. Data

We draw on a number of different data sources for the long-term analysis.

[6] Census *comarcas* are administrative areas within municipalities from 1995 national census that included as many as 10 small communities altogether totalling approximately 250 households. The marginality index was constructed from the census and included indicators of literacy, family size, and water and sanitation conditions.

[7] More generally, all CCT programs in Nicaragua were discontinued after 2006. In total they benefitted ~30,000 households. As such, general equilibrium effects on labor and marriage markets are likely to be limited.

Program Census Data—The 2000 program census provides baseline data on early and late treatment localities including household demographics, grades attained (defined throughout the paper as the highest grade attained, i.e. number of school grades passed) for all household members, housing characteristics, and assets.

Short-term Evaluation Surveys—The first round of the short-term evaluation surveys was conducted in September 2000, with subsequent rounds in 2002 and 2004. The household-level survey instrument was based on the 1998 Nicaraguan Living Standards Measurement Survey, with modules covering education, health, and household expenditures, among others. The sample was drawn from the 2000 program census and includes a random sample of early and late treatment households, 42 households in each of the 42 early and late treatment localities (Maluccio and Flores 2005). Attrition was approximately 10 percent per round. Starting in 2002, a non-experimental comparison group (that never received the program) was added from neighboring municipalities, chosen from a set of localities with marginality scores similar to the original 42 localities. The 2002 household survey targeted for interview 40 households in each of the 21 selected localities, and, since there is no program census data available for this comparison group, provides the baseline for this sample.

2010 Evaluation Survey—Between November 2009 and November 2011, 9 to 11 years after the start of the program, we conducted a long-term follow-up evaluation survey. Since the bulk of data were collected in 2010, and for convenience, we refer to this as the 2010 survey. We expanded the short-term household evaluation survey instrument, and added a separate, individual-level instrument with measures of learning and socio-emotional outcomes. The sample frame consists of all households in the original short-term randomized evaluation survey sample including the non-experimental comparison group, as well as an additional sample of households who, according to the 2000 program census, had children of critical ages relevant to the long-term evaluation. Specifically, we oversampled households with children born between January and June 1989 in the early and late treatment groups who were at risk of dropping out of school at the start of the program (see section IV for further details).[8] All experimental estimates account for this sampling procedure, using sampling weights, and also adjust for attrition as described in appendix F. While the detailed information from the earlier short-term evaluation

[8] We also oversampled households with children born between November 2000 and mid-May 2001, for analyses on exposure to the CCT during early childhood (Barham, Macours, and Maluccio 2013).

surveys is not available for the oversampled households, there is full information on them from the 2000 program census. The sample frame has a total of 1,330 households from the early treatment group, 1,379 households from the late treatment group, and another 757 households from the non-experimental comparison group in 2010.[9] For the 9–12 year old boys, the sample frame included 588 observations in the early treatment group, 550 in the late treatment group, and 263 in the non-experimental comparison group.

The 2010 survey was designed to collect data on households in the target sample, as well as all split-off households containing at least one of the original household members less than 22 years old in 2010. The household instrument included questions on educational attainment and current schooling, and a detailed survey instrument module to measure participation and earnings in all economic activities each household member had engaged in over the last 12 months. It asked separately about economic activities conducted at the place of residence and activities engaged in while temporarily absent from the household. For all young adults between 15 and 22, an additional module also collected their full labor market history, with questions on participation, location, and earnings for all non-agricultural wage jobs and self-employment. All questions in the household instrument were answered by the best-informed person available for the interview. Hence, answers were obtained from the young men themselves if they could by located at home, or from the household head or the spouse of the household head if not.

The 2010 individual instrument was conducted through direct in-person interviews with the respondents in their homes, and was designed to measure individual learning and socio-emotional outcomes of each individual born after January 1, 1988 (see appendix E for details). We administered three Spanish language and two math tests to all young adults between the ages of 15 and 22 years (in 2010). The Spanish language tests included a word identification test, a spelling test, and a test of reading fluency. The math tests included a test of math fluency and a test to measure problem solving at various levels of difficulty, which we refer to as math problems and similar to the *Peabody Individual Achievement Test* (Markwardt 1989). All tests were appropriate for different grade levels. In addition, we administered two tests that could capture both learning and cognitive development: the *Test de Vocabulario en Imágenes Peabody* (TVIP, the Spanish version of the *Peabody Picture Vocabulary Test*; Dunn et al. 1986), and a

[9] The 2002 household survey targeted 840 households in the comparison group, of which 823 were interviewed. Of those, 32 would have been ineligible for the CCT based on asset exclusion criteria (see appendix C). An additional 34 lived in border areas and ultimately received the program, so the target sample in 2010 was 757 households.

forward and backward digit span test, in which the respondent is asked to repeat a series of numbers read to him. Finally, the Raven's colored matrices (the 36-item version with sets A, AB, and B) was added to measure cognition (Raven, Court, and Raven 1984). Given that the intervention began when the cohort of boys was in late childhood, we did not necessarily expect large sustained program effects on cognition, and therefore included the Raven in the instrument to help separate more general cognitive skills from skills acquired in the classroom. An important advantage of all of the tests is that they provide observed, as opposed to self-reported, measures of learning and cognition, thereby substantially reducing concerns about reporting biases. Finally, the individual survey instrument also included measures of the socio-emotional outcomes of the young men. In particular, we implemented the 20-item Center for Epidemiologic Studies of Depression (CESD) Scale and the Strengths and Difficulties Questionnaire (SDQ).

Nicaraguan Population Census Data—We use the two most recent Nicaraguan censuses in the double difference models for the absolute effects. These national censuses provide repeated cross sections at the individual level and include basic demographic and education information, in 1995 before the start of the program and in 2005, the year the program ended.

IV. Methodology

Given the timing of the initial intervention and of subsequent data collection, no formal pre-analysis plan guides the analysis in this paper. Rather, the hypotheses and outcomes investigated broadly follow those outlined in proposals prepared to finance the 2010 survey data collection.[10]

A. Identification of Experimental Long-term Differential Impacts

To estimate long-term effects of the CCT on education, learning, and labor markets, we rely on the randomized phase-in to provide exogenous variation in program exposure. Since both experimental groups eventually received the program, we estimate differential treatment effects rather than absolute program effects. And so that the random variation provides enough statistical power to estimate these differential treatment effects, we focus on a cohort of boys whose educational attainment was more likely to have been affected by the program in the early than in the late treatment group. We do this by taking advantage of the fact that each treatment

[10] In particular for NSF (SES 11239945 and 1123993) and 3ie (OW2.216), and IDB funding. See https://sites.google.com/site/johnamaluccio/research/nicaragua-cct-proposal-and-analysis-plans-1.

group received the program for a different fixed three-year period, and exploiting pre-program school dropout patterns as well as program eligibility rules. This leads to a focus on the group of boys age 9-12 at the start of the program in our main analyses, though we also examine younger and older age cohorts covering the entire 7-13 eligible age range (Section V.D). More specifically, to determine a cohort of boys whose educational attainment was more likely to be influenced by the program in the early than in the late treatment groups, we consider several criteria, each of which point to slightly different age groups, but which together point to the 9-12 age cohort as the main cohort of interest.

First, we identify those who, in the absence of the program, were likely to have dropped out of school by the time it reached the late treatment group in 2003. Arguably, transfers provided at ages when most children are already enrolled in school and likely to continue, or after most children have already left school, will be less effective than transfers provided at ages when they are enrolled but at risk of dropping out (de Janvry and Sadoulet 2006).[11] Average pre-program enrollment rates by age at the start of the program (on November 1, 2000) are shown in Figure 1 (dashed line, left vertical-axis scale). After age 7, enrollment is at its highest (~0.7) until age 11, after which it declines indicating increasing risk of dropout for boys. Consequently, at some point between 2000 and 2003, prior to the late treatment group receiving the program, boys 9–12 years old in 2000 were enrolled but at high risk of dropping out. These boys would have been 12–15 by the time the program reached the late treatment group in January 2003, and thus more likely to already have dropped out of school, so that the program would likely affect their enrollment less.

Second, we consider the program rules and timing to determine the potential difference in exposure to the education transfers and conditionalities between treatment groups. Children between ages 7 and 13 at the start of the school year in January were eligible for the education transfer during that school year, provided they had not completed fourth grade. Abstracting from the grade restriction, we calculate the number of school years potentially affected by the education transfers and conditionalities for boys the same ages living in early and late treatment localities. Note that while both treatment groups received three years worth of transfers, because the early treatment group began receiving them in November 2000, i.e., *during* the school year,

[11] Children who would have been in school even in the absence of the transfer may still have been influenced in other ways, for example due to the requirement on number of days of attendance and the increase in household resources, some of which were designated for school uniforms and materials.

their transfers were potentially spread across as many as four school years. In contrast, the late treatment group began receiving their transfers in January 2003, spreading them across at most three school years. To determine the potential difference, one also needs to consider the child's age at the start of transfers in November 2000 as well as the age at the beginning of the school year in January when eligibility was determined. Comparing boys who were 11 years old at the start of the program, most in the early treatment would be eligible over (part of) four distinct school years (2000, 2001, 2002, and 2003), while most in the late treatment would be eligible for only one school year since they would turn 13 by January 2003.[12] This leads to a difference in exposure of three school years for 11-year olds, the age group we oversampled.[13] Similar comparisons make clear that the difference in exposure to the education transfer was three school years for both 11- and 12-year olds at the start of the program, two school years for 10- and 13-year olds, and so on.

Because we ignored the grade restriction, a concern with the above calculations is that they only crudely capture potential exposure differentials. In particular, (older) boys completing fourth grade prior to the beginning or end of the program in their localities would be ineligible in subsequent school years. And, while grades completed at baseline are observed, without strong assumptions they are insufficient to calculate ongoing eligibility for children in each school year, which depends on (endogenous) grade progression. But to gauge the potential relevance of the grade restriction, we can use program administrative data to examine *actual* exposure differentials. Figure 1 (black line, right vertical-axis scale) shows the differences between the average number of actual school years during which boys received transfers in the early treatment minus the average number for boys in the late treatment. Actual exposure differences are largest, approximately two school years, for the 10-year old boys and above 1.5 school years between ages 8 and 12.[14] Differences from the crude "expected" differentials calculated above

[12] For those 11-year olds whose birthdays are in November or December, however, they would only be eligible for three school years in early treatment and none in late treatment since they would be 14 years old by January 2003.
[13] The announcement of the program in July 2000 means that although the transfers did not begin until late in the school year 2000, the program still had strong potential to influence schooling in that year.
[14] Figure 1 shows the difference is higher than 1 at age 13 (the last year of eligibility), since the number represents the average for children who were 13 in November 2000, and many had not reached their 14th birthday (when they would become ineligible) until part way through the 2001 school year. There are also small positive differences for 14 year olds at start of program for similar reasons. Also, in the first year of the program, difficulties in correctly assessing ages led to some upward slippage and inclusion of some children older than 13.

stem from unobservable patterns of grade progression through to completed fourth grade during the program years, as well as non-compliance, both of which are endogenous.[15]

Based on the combined examination of risk of dropout and potential and actual exposure, we focus the analysis of the long-term differential effects on boys age 9–12 years at the start of the program. This is the age group for which we expect a large differential impact as a result of being randomly allocated into early versus late treatment. Although the actual differences in exposure are high for the 7 and 8 year olds, we do not include them in the main analyses because their risk of dropout is low during the program years. We also do not include the 13 year olds in the main analyses because many of them had already dropped out of school or completed fourth grade by program start. [16]

Analysis of the short-term intent-to-treat (ITT) estimates both before and after the phase-out of the early treatment group and the phase-in of the late treatment group further supports the choice of the 9–12-year old cohort for estimating long-term differential impacts (Table 1). In 2002 (top panel)—when the early treatment group had received transfers for two years but the late treatment group had not yet benefited—boys in the 9–12 year old age cohort had 0.36 additional grades attained, were 18 percentage points more likely to be enrolled in school, 36 percentage points more likely to attend school regularly (more than 85 percent of the time), and were 15 percentage points more likely to be literate (able to read and write according to parental report).[17] By 2004 (bottom panel), the transfers to the early treatment had ended and the late treatment group had started receiving transfers. In line with this change, the signs of the effects on enrollment and attendance in 2004 reverse, and differences between early and late treatment boys are no longer significant. Even so, the difference in grades attained between early and late treatment remains positive and significant (0.49 grades), and is one-third larger than in 2002.

The short-term results demonstrate that being exposed to the CCT during critical ages in primary school led to differences between the two experimental groups that persisted to 2004, after the early treatment group was no longer receiving transfers. Although not significant, at that point enrollment was higher in the late treatment group, who were still eligible for another year

[15] At the individual level, take-up of the education transfer during at least one school year was 88 percent for 9–12-year-old boys in the early treatment group.

[16] For the same reasons, and the substantial budgetary implications of tracking such a mobile population, we cut off tracking at age 12.

[17] These results build on prior evidence of the effect of the CCT program on education (Maluccio and Flores 2005; Maluccio, Murphy, and Regalia 2010), but focus on the specific age cohort relevant for analysis of the long term.

of program benefits. Consequently, the findings also suggest that estimates of the long-term differential effects of early versus late treatment may underestimate the absolute program effects.[18]

B. Empirical Specification for Long-term Differential Impacts

We estimate the following individual-level model,

$$Y_{il}^{k} = \alpha^{k} T_{l} + \boldsymbol{\beta^{k} X_{il}} + \varepsilon_{il}\,, k = 1 \ldots K \qquad (1)$$

where Y^k is one of the outcomes of interest for individual i in baseline locality l. T is an indicator that takes on the value of one for boys in localities randomly assigned to early treatment and zero for those in localities randomly assigned to late treatment. All analyses are carried out on an ITT basis and using all boys from both treatment groups in the 9–12-year-old age cohort, regardless of initial levels of completed schooling or actual program participation. Given randomized assignment, our main specifications limit the set of control variables **X** to the following: age when the program started in early treatment (dummy indicators for 3-month age groups); dummy indicators for whether the individual had 0, 1, 2, 3, or 4+ grades attained prior to the program; and regional fixed effects. All regressions also contain strata fixed effects to account for the fact that randomization was stratified by marginality level. All regressions are weighted to account for sampling and attrition providing population estimates (Section IV.C). We assess robustness of this specification with alternative sets of controls and using alternative weights and samples in appendix B. These alternative specifications do not substantively alter the principal findings.

Standard errors are adjusted for clustering at the locality level. As there are 42 localities, we also assess whether accidental imbalance related to the relatively small number of clusters drives any of the results. We follow Athey and Imbens (2017) and Young (2017) and estimate the exact p-value under the sharp null hypothesis that the treatment effect is null, by calculating all possible realizations of the test statistic and rejecting if the observed realization in the experiment itself is extreme enough. This avoids a dependence on asymptotic theorems that can produce inaccurate finite sample statistical inference sensitive to outliers. In testing the null of no treatment effects, randomization inference is not testing whether the average treatment effect is zero, but rather whether the treatment effect is zero for all participants.

[18] A second reason the experimental estimates may underestimate absolute treatment effects is positive spillovers between treatment groups. While randomization was done at the locality level, some of the localities share borders so that spillovers are possible. Appendix G demonstrates there is little systematic evidence of such spillovers.

C. Outcomes

CCTs can influence a wide range of behaviors and we therefore analyze a large set of outcomes, all measured in 2010 when the respondents are young men approximately 19–22 years old. To reduce concerns regarding multiple hypotheses testing we follow Kling, Liebman, and Katz (2007) and organize individual outcomes into different domains we refer to as "families of outcomes" capturing education, learning, labor market, earnings, and socio-emotional outcomes. For each individual outcome, we calculate within-sample z-scores, using the mean and standard deviation (SD) of the late treatment group. We then determine the average z-score for each family of outcomes and estimate the ITT effect using this index, which yields the effect size in SD. To further test the robustness of the findings for each family, we show p-values adjusted for familywise error rates, as proposed by Anderson (2008).

The education family includes an indicator of whether the boy was enrolled in school, grades attained, and whether he had completed grade 4, after which children were no longer subject to the program's schooling-related conditionalities. To analyze learning, we classify the tests into three categories. The first measures skills most likely learned in the classroom and comprises the average impact of the five achievement tests (word identification, spelling, reading and math fluency, and math problems). The second averages tests that are likely to capture both learning and cognition (receptive vocabulary and memory test). The third has only one test to proxy for cognition, the Raven's colored matrices. We refer to the three categories as the learning family, the learning and cognition family, and the cognition family, respectively.

For the labor market we consider two families of outcomes. The labor market participation family captures labor market participation and temporary migration for work. The earnings family includes labor market returns or earnings for work that is off the family farm. We present two versions of the earnings family to account for outliers. One uses absolute values of earnings trimmed at the top 5 percent of values, and the other uses the rank of earnings instead of the actual values. Labor market and earnings data were constructed based on a comprehensive module of labor market activity, covering all activities over the last 12 months, and separating between activities conducted while living in the home community versus those performed during periods of temporary migration. Additional variables are constructed from a separate labor market history module, which reflects *all* off-farm activities since the young men entered the

labor market. Given the high seasonality and temporary nature of many economic activities of the target population, this comprehensive approach is key to accurately reflecting labor market returns. Specific variables included in the earnings family are described in Section V.A.

For the socio-emotional outcomes, we conducted an exploratory factor analysis including all items of the CESD and the SDQ.[19] This analysis points to four latent factors, broadly capturing optimism, positive self-evaluation, stress, and negative self-evaluation (see appendix E for details). The family outcome is measured as the average of the z-scores for the four factors, after reversing the signs that have opposite meaning (stress and negative self-perception) so that higher values always indicate more positive, or better, socio-emotional outcomes.

D. Attrition and Internal Validity

Considerable effort and resources went into minimizing attrition in the 10-year follow-up for both the individual- and household-level instruments. Individuals who could not be found in their original locations were tracked to their new locations throughout Nicaragua. Migrants to Costa Rica—the destination of 95 percent of international migrants from the sample—also were tracked. As migration in this context is often temporary, multiple return visits by the survey team to the original localities were made to interview temporary migrants after they returned. In addition, for permanent migrants who could not be located, information on selected individual variables (related to educational, marital, and labor market status) was collected through proxy reports in the original household. For those select variables, information is available for all but 4 percent of the boys 9-12 years old at baseline. For the experimental sample used in the main analysis, only 10 percent of the boys who were 9-12 at baseline could not be tracked to their 2010 location and hence have missing information for the main outcomes coming from the household instrument, which importantly include labor market outcomes and earnings, as well as schooling, migration, and marital status.[20] Attrition is higher, 19 percent, for information obtained through the individual instrument, which required direct, in-person interviews and included all tests on learning, cognitive, and socio-emotional outcomes. Overall, these attrition levels are on par with or lower than those found in related longitudinal studies with a similar time horizons and target populations (see appendix F).

[19] We use exploratory factor analysis, as the correlations between items in the SDQ suggests standardized scoring of these items is unlikely to reflect intended latent traits, similar to findings in Laajaj and Macours (2017).
[20] In addition, 15 were deceased by 2010.

There are no significant differences in attrition levels between early and late treatment groups— differences are smaller than |1.5| percentage points and p-values of the differences are 0.598 for data from the household survey and 0.822 for the individual survey—and attrition does not affect the balance of baseline observables (appendix Tables A1, F1, and F2).[21] One approach would therefore be to introduce no further attrition corrections. However, more comprehensive analysis on attrition demonstrates that it is correlated with baseline characteristics that are associated with migration patterns, and that these correlations differ to some extent between early and late treatment groups. Consequently, even if attrition rates are relatively low, sample selectivity may still affect the findings.

To address this possibility, we consider several approaches to account for attrition selection, each relying on a different set of assumptions. In our preferred approach for the main analyses reported in the paper, we account for selection by weighting. Specifically, we use inverse probability weighting (IPW), allowing for differences between early and late treatment groups and incorporating information from the survey tracking process to give higher weight to individuals who were more difficult to find. The rationale underlying this strategy is that those not found are more similar on both observables and unobservables to those who were harder to find than to those more easily found (Molina Millán and Macours 2017). Separate weights are calculated for variables from the household and individual surveys. Appendix F provides further information on tracking and details the construction of these attrition weights, which also incorporate the sampling weights. Estimates without the attrition weights are shown in the appendix B (Tables B2.1-B2.4) and demonstrate that the main results are not driven by the attrition correction. We also estimate Lee (2009) bounds to test the sensitivity of results for corrections of the small differences in attrition rates, but note that their applicability to our study is unclear because the monotonicity assumption underlying them may be violated in this context in which we estimate differential treatment effects. Finally, as an additional robustness test we also use a more standard IPW correction. Unless otherwise noted, and consistent with the relatively low attrition rates for the main variables, the various approaches to correct for attrition yield qualitatively similar findings (see appendix B and results in Table B1.1).

[21] Attrition rates from the oversample households are not significantly different than for the rest of the sample and equally balanced (p-value of the differences are 0.510 for the household survey and 0.965 for the individual survey).

E. Identification of Non-Experimental Long-term Absolute Effects

Matching estimators using the non-experimental comparison group—To estimate the absolute effects on all outcomes, we take advantage of a non-experimental comparison group sample selected in 2002 and resurveyed in 2010. Since this group never received the CCT, we use it as a counterfactual for the early treatment group and compare 2010 outcomes of 9–12 year olds between the two groups.[22] While the non-experimental comparison localities were selected based on having similar marginality indices to the treatment localities and many share a geographical border, the ex ante match on locality-level characteristics did not fully balance household and individual characteristics across the groups. To improve the balance on observables, we estimate propensity scores at the individual-level and use the five nearest-neighbor matching estimator to estimate the absolute treatment effects. As the non-experimental estimates are based on much stronger assumptions than the experimental estimates, we present these results mainly as complementary evidence, and to facilitate interpretation of the differential effects.

The nearest neighbor estimator is bias corrected and standard errors are clustered at the locality level using the analytic asymptotic variance estimator developed by Abadie and Imbens (2008, 2011) that accounts for the fact that the propensity score is estimated. We use a typical "min-max" criteria to define the common support. Appendix H provides further details on the matching procedures, as well as a number of robustness tests, including specifications with alternative definitions of the common support, two nearest neighbors matching, and non-parametric kernel and local linear matching estimators.

Double difference estimation using 1995 and 2005 Nicaraguan National Censuses— At some point during the period 2000-05, the three-year CCT operated in *all* rural areas of the six municipalities where the evaluation was implemented. In addition to nearly the entire population of the 42 experimentally assigned localities, 80 percent of the population in the other 17 localities of these municipalities also received the program for three years, starting in late 2001 (appendix C). All together, therefore, the program covered over 90 percent of the rural population in these municipalities. Given this high coverage, it is possible to use national census data to approximate absolute program effects on selected outcomes.

[22] Attrition for this non-experimental sample is 15 percent for the variables measured in the household instrument and 29 percent for variables from the individual instrument. There are 223 boys in the comparison sample with education and labor market outcomes.

We use a non-experimental approach to estimate absolute program effects using a double difference model and the two most recent Nicaraguan censuses. Together, the censuses provide repeated cross sections at the individual level, in 1995 before the start of the program and in 2005, the year the program ended. The data include current municipality of residence (and whether rural or urban), as well as municipality of residence five years prior to the census administration date. As such, they allow us to account for selection due to domestic migration and therefore provide important complementary evidence that is less vulnerable to selection bias due to migration-related attrition compared to the 10-year follow-up survey. We calculate double difference impacts using two cohorts of boys (those ages 9–12 in 1990 and in 2000) by comparing education and marital status in rural areas of the six program municipalities to outcomes in rural areas of the six neighboring municipalities where the non-experimental comparison group was selected. The 9–12 age cohort in 2000 is the same age cohort examined in the experimental analyses. More specifically, we estimate

$$Y_{imt} = \delta_0 + \delta_1 T_{m,t-5} + \delta_2 C_t + \delta_3\, T_{m,t-5} * C_t + \varepsilon_{imt} \quad (2)$$

Where Y_{imt} is the outcome for boy i in municipality m measured in census year t, $T_{m,t-5}$ is an indicator for whether the boy resided in a treatment municipality five years prior to the census year, and C_t an indicator for the 2005 census. δ_3 yields the double difference estimate five years after the program began. Standard errors are robust to heteroskedasticity. We examine alternative age cohorts and comparison groups, and assess common trends in appendix I.

V. Results

A. Long-term Differential Impacts

In this section, we present the differential experimental ITT effects based on equation (1) for schooling, learning, labor market, earnings, and social-emotional outcomes. We present results for the family of outcomes as well as the components of the families. All tables include the mean of the late treatment group (and these are what are reported in the text below) for outcomes that are not measured in standard deviations. Robustness tests with alternative specifications and samples (section IV) are shown in appendix B and do not qualitatively alter the principal findings.

Schooling and Learning—Ten years after the start of the program in the early treatment group, schooling outcomes, as shown by the education family z-score, are significantly better in early compared to late treatment (Table 2, column 1). Columns 2-4 present the components of the education family; there is a significant differential effect on grades attained of nearly 0.3 grades (column 2), on an average of 5.5 grades attained. The significant differential effect estimated in 2004 (shown earlier in Table 1), largely persisted through 2010. Early treatment boys are also 4.5 percentage points (column 4) more likely still to be in school in 2010. Overall, a substantial minority, 18 percent in the late treatment group, is still in school, typically taking weekend classes so they can both work and go to school.[23] This pattern helps explain why average grades attained for the cohort increased substantially since 2004. Despite these gains, however, the overall level of schooling remains low, with only 75 percent having completed grade 4. These results for schooling demonstrate that boys exposed to the CCT during critical ages not only have higher grades attained but also are more likely to be continuing their studies, suggesting the grades attained differential is unlikely to diminish, and may increase further.

We next investigate the extent to which these differences in schooling are accompanied by learning gains and better labor market outcomes. In Table 3, we see that the improvements in schooling indeed translated into learning gains. Early treatment boys performed better than those in the late treatment on tests of math and Spanish. The difference is significant for each of the five individual tests administered (Table 4). This result is corroborated further by a significant increase in self-reported literacy (column 5, Table 2). Specifically, Table 3 column 1 shows that on average achievement test scores were 0.18 SD higher in the early treatment group, 10 years after the start of the program. Significant differential impacts are seen for the sorts of skills one would reasonably expect to be acquired in school: 0.16 SD for math (column 2) and 0.20 SD for Spanish reading and writing (column 3), but not for the more general cognitive outcomes.[24] In particular, the impact on the Raven in column 5, a test less likely to capture skills learned in the classroom, is close to zero (-0.02 SD). The mixed learning and cognition family that could capture both classroom and non-classroom skills (tests on receptive vocabulary and memory) show smaller and insignificant differentials in column 4, as do the individual tests that make up this measure (columns 6 and 7, Table 4). These results highlight that in Nicaragua, exposure to

[23] Of those enrolled in school, 15 percent are enrolled in tertiary education, while the rest are mainly in secondary school and a few still in primary, illustrating accumulated schooling delays.
[24] Achievement (math and Spanish) and cognitive tests were separated as specified in the project proposal.

the CCT during critical ages in primary school led to significant learning. Moreover, the magnitude of the differential effects are non-negligible and in line with absolute results from education interventions in other settings.

Labor market and earnings outcomes—Figure A1 presents the distributions of monthly off-farm earnings and shows a clear shift to the right for the early versus late treatment group. To contextualize this finding, consider a few key facts about the labor market experiences of this cohort (Table 5). First, virtually all individuals (98 percent) in the cohort (who in 2010 are young men 19–22 years old) are working, so that labor market effects cannot come from increased participation on the extensive margin. In fact, most young men combine different types of work—including work on the family farm (89 percent) with off the family farm (83 percent), referred to in the tables as worked off-farm. In addition, as noted earlier, 18 percent are studying while working.

Program effects on labor market participation show participation in off-farm work increased by 6 percentage points (Table 5, column 2). Given the low levels of schooling (less than 6 grades attained on average) it should not be surprising that almost all off-farm work is in non-skilled jobs, which is also where the increase is concentrated. Off-farm work includes work as agricultural laborers on farms not belonging to the household or on large plantations, salaried jobs in the non-farm sector (such as construction workers or security guards), and non-agricultural self-employment. Opportunities for remunerative off-farm work in such jobs are highly limited in the poor rural communities where the CCT operated. Table 5 therefore presents the differential impact on seasonal migration for work. Boys in the early treatment are 9 percentage points more likely to have migrated temporarily for work in the last 12 months (column 3), nearly one-third higher than the average (31 percent).[25] Similarly, columns 4 and 5 highlight that young men in early treatment are 8 percentage points more likely to have held a non-agricultural wage job at some point, and 7 percentage points more likely to have held an urban wage job (a 50 percent increase over the mean). Taken together as a family capturing labor market participation, these different indicators point to a substantial and significant 0.27 SD shift (column 1) in the type of economic activities these young men engage in and an increased degree

[25] This is not due to migration starting earlier, as there were no significant differences between early and late treatment group in work or migration patterns for this cohort in 2004.

of temporary mobility. On the other hand, the program effect on permanent migration out of the municipality, shown in column 6, is close to zero.

In Table 6, we explore whether the changes in labor market participation and seasonal migration are accompanied by differential effects on earnings, as suggested by Figure A1.[26] Earnings measures include average and maximum monthly earnings, as well as total earnings in the last 12 months (columns 2–4). The final column shows the monthly salary in the highest paid non-agricultural salary job ever held, including prior to the previous 12 months. To avoid selectivity bias, all values are unconditional with zero earnings when an individual did not have off-farm (or salary) earnings during the reference period.[27] As the earnings data are highly skewed, we follow Athey and Imbens (2017) and estimate the effect on the ranks of each of the indicators in panel A. Alternatively, we trim the 5 percent highest outliers in panel B. Results are broadly consistent across the different indicators and methods, with the earnings family outcome in column 1 showing an overall increase of about 0.2 SD for both the rank and the absolute value in the earnings family. The point estimate for average monthly earnings in column 2 indicates a differential increase of 14 percent, with estimates of the other indicators showing differential increases between 7 and 62 percent. Finally, Table 7 shows the results of a quantile regression for the earnings family. The results suggest positive gains across the distribution, with point estimates that tend to increase for higher deciles. Estimates are mostly significant between the 40th and the 80th percentiles, and gradually increase from 0.15 SD to 0.3 SD in this range.

The findings highlight that differential exposure to the CCT during primary school led to non-negligible increased returns in the labor market for young men in Nicaragua. Moreover, the results are likely underestimates of the long-term annual returns for three reasons: 1) they capture only the differential effects; 2) young men in the early treatment group are still more likely to be

[26] We do not include earnings on the family farm, as person-specific individual returns cannot be quantified in this setting. Only 11 percent of this cohort is head of their own household and almost all work on the farm of an older household member, typically their father or father-in-law. Moreover, we do not have information on agricultural inputs, preventing us from calculating agricultural profits for household farms. While in theory it is possible that the higher earnings we measure off-farm are offset by lower earnings on-farm, the ITT differential impacts on participating in on-farm activities are in fact positive (though small and insignificant), suggesting that at least on the extensive margin off-farm work does not substitute for on-farm work. Additionally, while total months worked off-farm did not change (and in fact the point estimate is negative), time worked during seasonal migration increases (from 30 to 40 days on average) while the time worked off-farm in the village of origin decreases.

[27] As relatively few men have ever had a salaried job, the unconditional mean of the monthly salary is much lower than for the off-farm earnings in the last 12 months.

enrolled in school perhaps leading to higher earnings when they complete their schooling; and 3) these returns are measured early in the working life cycle.

Socio-emotional outcomes—Last, we examine the effects on socio-emotional outcomes as soft skills are another pathway through which the program could affect labor market outcomes. It is also possible that labor market or other outcomes in turn affect these measures. Table 8, column 1, shows a small (0.05 SD) and insignificant differential improvement in socio-emotional outcomes. Considering each of the four factors separately, however, reveals that the average masks offsetting effects on the different latent traits. The differential effects on positive self-evaluation and optimism are positive, significant, and relatively large at around 0.25 SD (columns 2 and 3), possibly reflecting higher learning and earnings for the early treatment group. At the same time, however, the early treatment group exhibits more stress, and more likely to agree with statements referring to negative self-evaluation, even if the point estimates in columns 4 and 5 for these outcomes are smaller, around 0.16 SD. A possible explanation for this latter finding could be higher stress related to temporary migration.

Robustness—Appendix B (tables B1.1 and B1.2) shows that the education, learning, labor market, and earnings results are generally robust to various alternative definitions of weights or sample, to different assumptions related to attrition, to adjustments for multiple hypothesis testing (using Anderson's familywise error rate) and to randomization inference.

B. Interpretation

The findings provide positive evidence regarding the long-term returns to CCTs and their potential to reduce the intergenerational transmission of poverty, in contrast with some other studies on CCTs that suggest limited learning and labor market returns. Setting aside the methodological differences discussed in the introduction, we further explore why, in this context, the program appears to have delivered on its promise of more learning and earnings.

One potentially important difference with other settings is that the population studied in this paper is poorer and starts from much lower levels of education than many of the previously studied CCTs. The cohort studied in this paper had only 1.2 grades attained at the start of the program when they were 9-12 years old and more than half of their household heads had no education. As a result, it is possible that additional time in the classroom as a result of the CCT led to learning gains more effectively in this context than in settings where children had already

acquired basic language and math skills prior to the introduction of the CCT. If this interpretation is correct, one could expect impacts to be larger for the poorest segments of the target population, for whom education levels are even lower.

We take advantage of the stratification of the randomized experiment to investigate such treatment heterogeneity, distinguishing between children from the three poorest strata and the four other better off strata. Results in Table 9 confirm that the ITT effects on learning are indeed larger for boys in the poorest strata compared to the better off strata. Differential education and learning gains are 0.67 SD and 0.36 SD, respectively, for the three poorest strata but close to zero for the better off strata, suggesting that the differential learning results are driven primarily by the poorest half of the population who had lower average levels of education at baseline.[28] Hence encouraging children to enroll and stay in school may have particularly high returns for learning among the poorest, possibly because the marginal gains are higher for them than for children in the better off strata.[29]

Given the overall low levels of education, it is not surprising that improvements in these basic skills did not automatically lead to skilled wage jobs. Findings from qualitative interviews in 2009, however, suggest that such skills may have led to higher labor market returns in seasonal migration (CIERUNIC 2009). While temporary migration has costs, there are important differences in wages across regions in Nicaragua and wages are substantially higher in Costa Rica. Better math skills may have helped the young men assess the cost-benefit trade-offs of seasonal migration, and better reading and writing skills may have enabled them to complete any relevant paper work, particularly for international migration. That said, Table 9 does not show significant differences in labor market returns between the poorer and better off strata. This may indicate that earning gains are not necessarily the direct result of additional grades attained alone.

The gains in off-farm earnings and the role of temporary migration in improving earnings could also reflect other direct or indirect effects of the CCT program.[30] One possible additional mechanism for the migration and earnings results could be that the CCT increased the strength of local networks (e.g., by children spending more time in school together or beneficiaries attending

[28] At baseline children in the three poorest strata on average had 1 grade attained, versus 1.5 in the better off strata.
[29] The cross-sectional relationship between learning and grades attained also shows decreasing returns to additional grades, with the slope coefficient steepest for children with less than third grade.
[30] As there is no overall differential impact on the cohort's fertility (Table A4) and given that reproductive decisions for young men in any case do not have strong labor market implications in this context, a possible fertility channel does not seem likely.

regular meetings) and that this in turn led to lower costs or improved information about different opportunities. Variation in pre-program family networks and geographical variation in the density of the early treatment localities suggest some of these additional mechanisms may indeed be at play, but lack statistical power to make conclusive inferences (appendix G).[31]

More generally, we do not interpret the results as the effects of the education transfers and schooling conditions alone, but rather as the result of having been exposed to the entire package offered by the CCT program. Households were also eligible to receive the nutrition and health transfer, independent of whether their children were eligible for the education transfer. And, children who did not meet the eligibility requirements of the education transfers often lived in households with younger children who would have been eligible for them. By relaxing the household liquidity constraint, the different program transfers might have enabled children to stay in school even beyond the ages and grade levels the education transfers were targeting.

C. Non-experimental Absolute Effects

Matching—In Table 10, we present non-experimental matching estimates of the absolute program effects using the five nearest neighbors (NN5) bias-adjusted estimator to compare 2010 outcomes between the early treatment group and the non-experimental comparison group (see appendix H for more details on methods and sensitivity analyses). The findings confirm the overall direction and interpretation of the experimental differential results. The estimates indicate an increase of 1.4 in grades attained and 0.4 SD in learning, 10 years after the start of the program or 7 years after the program ended in the early treatment group. These results are consistent with the pattern of differential results, which not only showed a significant differential in grades attained after 10 years, but also in enrollment. We interpret the difference in the absolute and differential program effects as evidence that exposure to the CCT increased school attainment and learning for both the early and late treatment groups, not only because it kept children in school during the program, but also because households (or the children themselves)

[31] Another possible spillover includes intra-household spillovers in the late treatment group that could possibly lead to an overestimate of the treatment effect if late treatment households reallocated resources away from the 9-12 year olds towards younger children who would become eligible after the program started in their localities. There are too few children (less than 17%) without younger siblings to consider heterogeneity along these dimensions. That said, as the timing of the start of the program in the late treatment was not announced at the outset and the data suggest an overall increase in enrollment patterns in the late treatment even prior to them receiving the program (a possible anticipation effect), this seems unlikely to have played a major role. If anything, the apparent anticipation effect on enrollment in the late treatment group would suggest that the differential ITT estimates are underestimates.

continued to invest in education after the program ended. Results for the other outcome families overall point in the same direction as the experimental differential results, but are not significant. But given the relatively large standard errors, they do not rule out positive absolute effects of the same size or larger than the differential findings.

Double Difference—To provide additional evidence of absolute program impacts with a different identification strategy, Table 11 presents double difference results that compare treatment municipalities to comparison municipalities using the Nicaraguan national census. The double difference results show improvements for the four educational indicators in 2005, all of which are similar or larger than the 2010 experimental differential effects. By 2005, boys in the 9–12 cohort (who would have been 14–17 in 2005) in treatment municipalities had attained 0.6 more grades than similarly aged boys in comparison municipalities, were 9 percentage points more likely to be literate, and were nearly 4 percentage points more likely to be enrolled.

These double difference absolute effects may be smaller than the 10-year absolute effects because: 1) they are measured after a shorter period (2005 compared to 2010); 2) the program was implemented at different times in different areas and was still operating in the late treatment area in 2005; and 3) there may have been continued investment in children's education even after the end of the program.

D. Differential and Absolute Program Effects for Other Age Groups

Per the CCT program rules, the 7-8 and 13 year old cohorts were potentially eligible for the education transfers. We therefore also analyze whether long-term program impacts can be detected for these other age groups.

The experimental phase-in design is not well suited to estimating long-term program effects for the younger age 7-8 cohort, as the potential difference in exposure between early and late treatment group in this cohort is limited (both groups could have received the education transfers for the full program period in their respective treatment groups if they did not complete fourth grade). As a result, we expect the differential program effects to be smaller than for the 9-12 year old cohort. Table A2 confirms that the differential ITT effects are small (panel A), and for the education and learning families close to zero. Even so, with the exception of education, panel B confirms that the findings for the 9-12 year old cohort are robust to widening the age group to include the 7 and 8 year olds.

The lack of differential effects does not necessarily mean, of course, that the younger cohorts did not benefit from the program in absolute terms. We explore this possibility by estimating the non-experimental double difference impacts using the national census data (Table I1). These estimates show large absolute increases in the available educational outcomes for the 7-8 year olds that are statistically indistinguishable from the older (9-12) cohort. Table I1 also demonstrates there were significant double difference effects on most of the educational outcomes for the 13 year-old cohort by 2005.

The 2010 survey did not track 13 year olds beyond their original households nor administer the individual-level instrument to them. Therefore, we are unable to estimate experimental differential results for all of the same outcomes as for the younger age cohorts. However, proxy information for a number of key outcomes was collected, and in some cases the respondent was tracked and interviewed in the split-off household of a younger sibling. Including direct and proxy reports, data on those outcomes is available for approximately 94 percent of the boys in the 7-13, 9-13, and 13 year old age groups and rates for missing information are balanced across early and late treatment groups.

Recognizing that the proxy information is likely to have more measurement error than direct reports, we first re-estimate differential effects for the 9-12 year olds on the available outcomes including the proxy information (Table A3, Panel A). Results are similar to the main findings, though less precise.[32] Results for the combined 9-13 year old cohort (Panel B) are similar in magnitude and significance to the 9-12 year old cohort alone. Moreover, there are no statistically significant differences between the 9-12 versus 13-year olds for any of the outcomes (Panel C). Last, analysis of the entire 7-13 year old age cohort incorporating proxy information is presented in Panel D and confirms that while including the youngest cohort reduces the magnitudes of the differential impacts (in line with the findings discussed above), the program effects remain statistically significant at the 10 percent level or below for all but participation in off-farm work.

E. Cost-Benefit Analysis

Increased earnings for the 9-12 age cohort suggest there may be potential for recouping costs of the CCT. A full cost-benefit analysis would require monetary assessment of all of the resource

[32] The analysis incorporating the proxy information also provides an additional test of possible attrition bias of the main results for the 9-12 year age group, since the overall attrition rates are substantially lower. This further suggests that overall findings do not appear to be driven by attrition bias.

costs as well as all of the potential (positive or negative) effects of the CCT. Assessment is further complicated because we can only estimate differential effects with the experiment and do not have experimental evidence on absolute effects, although the evidence strongly suggests for outcomes considered in this paper they are at least non-negative. With these caveats, and under fairly conservative assumptions regarding cost and benefit flows, using a discount rate of 10 percent, the net present value (NPV) turns positive around 2027. Appendix J provides the details.

VI. Conclusions

CCTs and related interventions—combining short-term poverty reduction with enhanced investment in human capital to strengthen the productive capacity of future generations—have broad policy appeal. Numerous rigorous empirical studies have established that such programs have been effective at reducing contemporaneous poverty and increasing nutrition, health, and children's school attainment in the short term. Establishing and understanding the longer-term impacts of these programs is necessary for assessing the extent to which the investments in human capital fulfill their promise of improving the welfare of the next generation.

Building on the short-term randomized evaluation of a CCT program for which sizable short-term impacts on educational outcomes had been demonstrated earlier, we designed a 10-year follow-up study addressing the most important pitfalls for long-term analysis of human capital and labor market outcomes. In particular, we tracked migrants domestically and internationally to reduce attrition, and implemented the survey at a point when entry into the labor market for the cohort of interest was complete. Using the experimental variation in exposure, we estimate differential program effects between early and late treatment groups.

The experimental results indicate the CCT led to long-term gains in schooling, learning, and labor market returns. The increased labor market returns reflect employment gains through temporary migration. One plausible interpretation is that with more education and learning, the young men developed core competencies that made them better at finding higher paying jobs that are often further away from home. Finally, while virtually all are working, because the impacts are found at the start of these young men's working lives, they potentially set them on a higher earnings trajectory for the future.

The sustained learning improvements are important in their own right, and relate to the global sustainable development goal of quality education. In addition, we find that impacts on learning are larger for the poorest households. This has implications for targeting of CCTs, which currently do not always reach the poorest populations, even in countries with widespread coverage. Equally important, while the gains are substantial, the population started with low initial levels education and face many other constraints. It is unsurprising, therefore, that the program did not completely transform the lives of these young men and their families to a fundamentally higher level of welfare.

Overall, our results show that exposure to a CCT during critical ages in primary school translated into gains in schooling, learning, and earnings, and point to important and sustained benefits several years after the end of the program. Based on the earnings gains, and under relatively conservative assumptions regarding cost and benefit flows for this subset of beneficiaries, the program achieves positive NPV within two decades. Despite skepticism in some policy circles, the evidence we present, alongside a few other recent studies, suggest these widespread interventions may well have a role to play in reducing the intergenerational transmission of poverty.

VII. References

Abadie, A., and G. Imbens. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76(6): 1537–1557.

Abadie, A., and G. Imbens. 2011. "Bias-Corrected Matching Estimators for Average Treatment Effects." *Journal of Business & Economics Statistics* 29(1): 1–11.

Aizer, A., S. Eli, J. Ferrie, and A. Lleras-Muney. 2016. "The Long-Run Impact of Cash Transfers to Poor Families." *American Economic Review* 106(4): 935–971.

Akee, R., W. Copeland, G. Keeler, A. Angold, and E.J. Costello. 2010. "Parent's Incomes and Children's Outcomes: A Quasi-Experiment with Casinos on American Indian Reservations." *American Economics Journal: Applied Economics* 2(1): 86–115.

Anderson, M. 2008. "Mutiple Inference and Gender Difference in the Effects of Early Interventions: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of American Statistical Association* 103(484):1481-1495.

Araujo, M. C., M. Bosch, and N. Schady. 2016. "Can Cash Transfers Help Households Escape an Intergenerational Poverty Trap?" *NBER Working Paper* No. 22670.

Athey, S., and G.W. Imbens. 2017. "The Econometrics of Randomized Experiments." Banerjee, A. and E. Duflo, (eds.), *Handbook of Economic Field Experiments*. Volume 1. Elsevier.

Baez, J. E., and A. Camacho. 2011. "Assessing the Long-Term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia." *World Bank Policy Research Working Paper* No. 5681. Washington, DC, United States: World Bank.

Baird, S., C. McIntosh, and B. Özler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly Journal of Economics* 126(4): 1709–1753.

Baird, S., C. McIntosh, and B. Özler. 2016. "When the Money Runs Out: Do Cash Transfers have Sustained Effects?" *World Bank Policy Research Working Paper* No. 7901.

Baird, S., J.H. Hicks, M. Kremer, and E. Miguel. 2016. "Worms at Work: Long-run Impacts of a Child Health Investment." *Quarterly Journal of Economics* 131 (4): 1637–1680.

Bandiera, O. R. Burgess, S. Gulesci, N.Das, I. Rasul, M. Sulaiman. 2017. "Labor Markets and Poverty in Village Economies." *Quarterly Journal of Economics*, 132(2):811-870.

Banerjee, A., E. Duflo, R. Chattopadhyay, and J. Shapiro. 2016. "The Long Term Impacts of a "Graduation" Program: Evidence from West Bengal." Unpublished, MIT.

Barham, T., K. Macours, and J.A. Maluccio. 2013. "Boys' Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions," *American Economic Review: Papers and Proceedings* 103(3): 467–471.

Barham, T., K. Macours, and J.A. Maluccio. 2018. "Experimental Evidence of Exposure to a Conditional Cash Transfer During Early Teenage Years: Young Women's Fertility and Labor Market Outcomes" *CEPR Discussion Paper* 13165.

Barrera-Osorio, F., Linden, L.L., and Saavedra, J.E. (2018). "Medium- and Long-Term Educational Consequences of Alternative Conditional Cash Transfer Designs: Experimental Evidence from Colombia." *American Economic Journal: Applied Economics,* forthcoming.

Behrman, J. R., S.W. Parker and P.E. Todd. 2009a. "Medium-Term Impacts of the *Oportunidades* Conditional Cash Transfer Program on Rural Youth in Mexico." In: S. Klasen and F. Nowak-Lehmann, eds. *Poverty, Inequality, and Policy in Latin America*, 219–270. Cambridge, United States: MIT Press.

Behrman, J. R., S.W. Parker and P E. Todd. 2009b. “Schooling Impacts of Conditional Cash Transfers on Young Children: Evidence from Mexico.” *Economic Development and Cultural Change* 57(3): 439–477.

Behrman, J. R., S.W. Parker and P.E. Todd. 2011. “Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Followup of PROGRESA/Oportunidades.” *Journal of Human Resources* 46(1): 93–122.

Bettinger, E., M. Kremer, M. Kugler, C. Medina, C. Posso, and J.E. Saavedra. 2016. “Can Educational Voucher Programs Pay for Themselves?” Unpublished, Stanford.

Centro de Investigación y Estudios Rurales y Urbanos de Nicaragua (CIERUNIC). 2009. “Qualitative Findings for the Evaluation of the Long-term Impact of the *Red de Protección Social* in Nicaragua.” Unpublished.

Cesarini, D. E. Lindqvist, Ötline, B. Wallace. 2016. “Wealth, Health, and Child Development: Evidene from Administrative Data on Swedish Lottery Players.” *Quarterly Journal of Economics* 131(2):687–738.

de Janvry, A., and E. Sadoulet. 2006. “Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality.” *World Bank Economic Review* 20(1): 1–29.

Duflo, E., P. Dupas, and M. Kremer. 2015. “Education, HIV, and Early Fertility: Experimental Evidence from Kenya.” *American Economic Review* 105(9): 2757–2797.

Dunn, Lloyd M., D.E. Lugo, E.R. Padilla, and Leota M. Dunn. 1986. *Test de Vocabulario en Imágenes Peabody*. Circle Pines, Minnesota: American Guidance Service, Inc.

Duque, V., Rosales, M. F., and Sanchez, F., 2018. “How Do Early-Life Shocks Interact with Subsequent Human-Capital Investments? Evidence from Administrative Data.” Mimeo, University of Michigan.

Fernald, L., P. Gertler and L. Neufeld. 2009. “10-Year Effect of Oportunidades, Mexico's Conditional Cash Transfer Programme, on Child Growth, Cognition, Language, and Behaviour: a Longitudinal Follow-up Study.” *Lancet* 371:828–837.

Filmer, D., and N. Schady. 2014. “The Medium-Term Effects of Scholarships in a Low-Income Country.” *Journal of Human Resources* 49(3): 663–694.

Fiszbein, A., and N. Schady. 2009. “Conditional Cash Transfers: Reducing Present and Future Poverty.” *World Bank Policy Research Report*. Washington, DC: World Bank.

García, A., Romero, O.L., Attanasio, O. and Pellerano, L. 2012. “Impactos de Largo Plazo del Programa Familias en Acción en Municipios de Menos de 100 mil Habitantes en los Aspectos Claves del Desarrollo del Capital Humano.” Technical report, Union Temporal Econometria S.A. SEI. con la asesoria del IFS.

Gertler, P. J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S.M. Chang, and S. Grantham-McGregor. 2014. “Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica.” *Science* 344(6187) 997–1001.

Gertler, P., S. Martinez and M. Rubio-Codina. 2012. "Investing Cash Transfers to Raise Long Term Living Standards." *American Economic Journal: Applied Economics* 4(1):164–192.

Haushofer, J. and J. Shapiro. 2018. “The Long-Term Impact of Unconditional Cash Transfers: Experimental Evidence from Kenya.” Unpublished, Princeton.

Hoddinott, J., J.A. Maluccio, J.R. Behrman, R. Flores, and R. Martorell. 2008. “Effect of a Nutrition Intervention During Early Childhood on Economic Productivity in Guatemalan Adults.” *Lancet* 371(9610, 2): 411–416.

Hoynes, H., D.W. Schanzenbach, and D. Almond. 2016. "Long-Run Impacts of Childhood Access to the Safety Net." *American Economic Review* 106(4):903–934.

Kling, J., J. Liebman and L. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1): 83–119.

Laajaj, R. and K. Macours. 2017. "Measuring Skills in Developing Countries." *World Bank Policy Research Paper* No. 8000. Washington, DC: World Bank

Lee, D.S. 2009. "Training, Wagers, and Sample Selection: Estimating Sharp Bounds on Treatment Effect." *The Review of Economic* Studies 76(3):1071–1102.

Levy, S., and N. Schady. 2013. "Latin America's Social Policy Challenge: Education, Social Insurance, Redistribution." *Journal of Economic Perspectives* 27(2): 193–218.

Løken, K. V., M. Mogstad and M. Wiswall, 2012. "What Linear Estimators Miss: The Effect of Family Income on Child Outcomes." *American Economic Journal: Applied Economics*, 4, 1-35

Macours, K. and R. Vakis. 2016. "Sustaining Impacts When Transfers End: Women Leaders, Aspirations, and Investment in Children." *NBER Working Paper* No. 22871.

Macours, K., N. Schady and R. Vakis. 2012. "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment." *American Economic Journal: Applied Economics* (4)2: 247–273.

Maluccio, J. A., A. Murphy and F. Regalia. 2010. "Does Supply Matter? Initial Schooling Conditions and the Effectiveness of Conditional Cash Transfers for Grade Progression in Nicaragua." *The Journal of Development Effectiveness* 2(1): 87–116.

Maluccio, J. A., and R. Flores. 2005. "Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan *Red de Protección Social.*" *Research Report 141*. Washington, DC, United States: International Food Policy Research Institute.

Markwardt, F. C. 1989. *Peabody Individual Achievement Test-Revised Manual*. Circle Pines, Minnesota, United States: American Guidance Service.

Molina Millán, T. and K. Macours. 2017. "Attrition in Randomized Control Trials: Using Tracking Information to Correct Bias." *CEPR Discussion Paper No. DP11962*

Molina Millán, T., K. Macours, J.A. Maluccio, and L. Tejerina. 2018b. "Experimental Long-term Effects of Early Childhood and School-age Exposure to a Conditional Cash Transfer Program." Nova University, unpublished.

Molina Millán, T., T. Barham, K. Macours, J.A. Maluccio, and M. Stampini. 2018a. "Long-term Impacts of Conditional Cash Transfers: Review of the Evidence." *World Bank Research Observer*, forthcoming.

Parker, S. W and T. Vogl. 2018. "Do Conditional Cash Transfers Improve Economic Outcomes in the Next Generation? Evidence from Mexico." *NBER Working Paper* No. 24303.

Raven, J.C., Court, J.H. and Raven, J. 1984. *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2: Coloured Progressive Matrices*. London: H. K. Lewis.

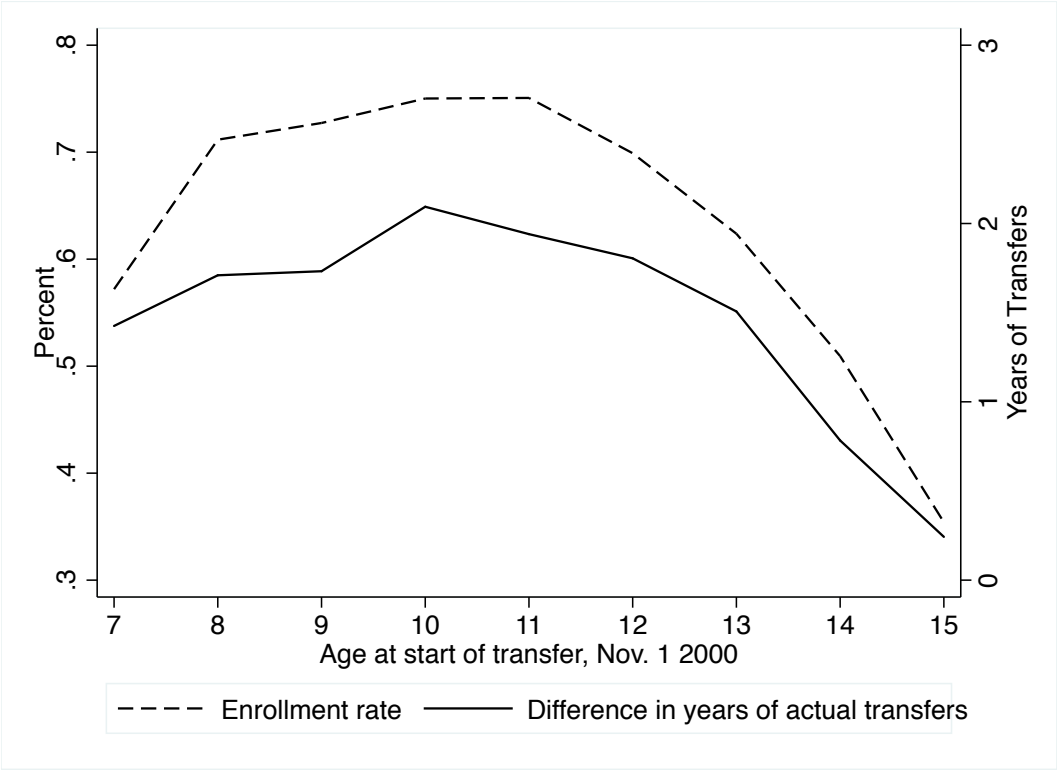
Robles, M., M.G. Rubio, and M. Stampini. 2018. "Have Cash Transfers Succeeded in Reaching the Poor in Latin America and the Caribbean?" *Development Policy Review*, forthcoming.

World Bank. 2015. *The State of Social Safety Nets 2015*. Washington, DC: World Bank.

Young, A., 2017. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results", LSE, unpublished

Figures and Tables

FIGURE 1: DIFFERENCE IN MEAN YEARS OF EDUCATION TRANSFERS RECEIVED BETWEEN EARLY AND LATE TREATMENT GROUPS AND MEAN ENROLLMENT RATE FOR BOYS



Notes: Source: Enrollment rates (left vertical axis scale) are calculated using the 2000 program census. The difference in years of education transfers are calculated using program administrative data (right vertical-axis scale). The program administrative data contain information on education transfers between 2000 and 2005. The difference in mean years of education transfers refers to the mean difference in the total number of school years that all children in the early and late treatment localities received, not just boys in the sample. Since eligibility depended on age at the start of the school year, while we calculate ages based on the start of the program in the early treatment group (November 1, 2000), it is possible for children age 13 to have received more than one year of transfers as described in the text.

TABLE 1: 2002 AND 2004 EXPERIMENTAL IMPACTS ON EDUCATION

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Attended School More Than 85% of Time =1	Read and Write =1
	(1)	(2)	(3)	(4)	(5)
Panel A: 2002 — Absolute Effects					
ITT	0.361***	0.053*	0.182***	0.360***	0.150***
	(0.094)	(0.031)	(0.042)	(0.055)	(0.034)
N	475	475	475	475	475
R^2	0.828	0.747	0.191	0.271	0.324
Mean late treatment	2.396	0.277	0.733	0.544	0.735
Panel B: 2004 — Differential Effects					
ITT	0.487***	0.086*	-0.049	-0.100	0.124***
	(0.155)	(0.045)	(0.063)	(0.066)	(0.029)
N	458	458	458	458	458
R^2	0.598	0.467	0.262	0.239	0.241
Mean late treatment	3.585	0.536	0.626	0.564	0.815

*Notes:*** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Compares ITT effects of early versus late treatment groups. The late treatment group started to receive the program in 2003, so 2002 represents absolute effects and 2004 differential effects. Results for boys that were 9–12 in 2000. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Attended school for more than 85% of the time is zero for those who were not enrolled in school at the time. All variables measured using the 2002 and 2004 household instrument.

TABLE 2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EDUCATION FAMILY AND LITERACY

	Education Family Z-Score	Education Family Components			Read and Write =1
		Grades Attained	Completed Grade 4 =1	Enrolled =1	
	(1)	(2)	(3)	(4)	(5)
ITT	0.098**	0.288*	0.035	0.045**	0.052**
	(0.043)	(0.167)	(0.024)	(0.021)	(0.021)
N	1,007	1,006	1,006	1,005	1007
R^2	0.346	0.425	0.356	0.096	0.175
Mean late treatment	-0.026	5.498	0.747	0.181	0.874

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score averages the z-scores for the individual components and is calculated if any component is available. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household instrument.

TABLE 3: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LEARNING AND COGNITION FAMILIES (Z-SCORES)

	Learning			Mixed Cognition and Learning	Cognition (Raven)
	Math and Spanish	Math	Spanish		
	(1)	(2)	(3)	(4)	(5)
ITT	0.183**	0.160**	0.204**	0.113	-0.016
	(0.070)	(0.069)	(0.081)	(0.082)	(0.095)
N	907	905	907	906	906
R^2	0.448	0.395	0.437	0.380	0.215

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 individual instrument.

TABLE 4: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LEARNING AND COGNITION BY TEST (Z-SCORES)

	Learning Family Components					Mixed Cognition and Learning Family Components	
	Math Fluency	Math Problems	Reading Fluency	Spelling	Word Identification	Receptive Vocabulary	Memory Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ITT	0.183**	0.179**	0.137**	0.252***	0.206**	0.128	0.094
	(0.070)	(0.082)	(0.067)	(0.078)	(0.087)	(0.102)	(0.074)
N	907	904	904	898	905	906	902
R^2	0.448	0.349	0.346	0.425	0.358	0.353	0.291

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 individual instrument.

TABLE 5: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LABOR MARKET PARTICIPATION, AND MIGRATION

	Labor Market Participation Family Z-Score	Labor Market Participation Family Components				Permanent Migration Out of Municipality =1
		Worked Off-Farm =1 (last 12 months)	Migrated for Work =1 (last 12 months)	Ever Had a Salaried Non-Agricultural Job =1	Ever Worked in Urban Area =1	
	(1)	(2)	(3)	(4)	(5)	(6)
ITT	0.272***	0.062***	0.093***	0.084**	0.065*	-0.019
	(0.075)	(0.022)	(0.032)	(0.036)	(0.034)	(0.028)
N	1,006	1,006	1,006	998	998	1,007
R^2	0.146	0.074	0.169	0.132	0.115	0.102
Mean late treatment	-0.018	0.828	0.312	0.226	0.127	0.150

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household instrument.

TABLE 6: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EARNINGS FAMILY AND COMPONENTS

	Family Z-Score	Earnings Family Components (C$)			
		Earnings Per Month Worked (last 12 months)	Annual Earning (last 12 months)	Maximum Earnings (last 12 months)	Maximum Salary Ever
	(1)	(2)	(3)	(4)	(5)
Panel A: Rank of Earnings					
ITT	0.194***	41.780**	25.568	49.313**	43.899**
	(0.057)	(19.471)	(18.493)	(19.684)	(19.290)
N	1,006	1,006	1,006	1,006	998
R^2	0.097	0.082	0.095	0.094	0.094
Mean late treatment		497	503	498	486.9
Panel B: Earnings — Five Percent Trim					
ITT	0.192***	201.152***	595.013	211.421***	142.260*
	(0.067)	(63.624)	(619.322)	(69.318)	(71.919)
N	997	956	956	956	955
R^2	0.097	0.085	0.084	0.107	0.071
Mean late treatment		1436	8222	1619	228

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Earnings include wage work off the family farm. Earnings in panel A are trimmed at the top five percent of values. Earnings are in Nicaragua Cordobas (C$) and the exchange rate is approximately 20. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household instrument.

TABLE 7: 2010 DIFFERENTIAL QUANTILE REGRESSIONS ON EARNINGS FAMILY (Z-SCORE)

	Percentile of Earnings Family								
	10	20	30	40	50	60	70	80	90
ITT	0.219	0.130	0.133	0.150*	0.172*	0.230*	0.223	0.300**	0.154
	(0.153)	(0.134)	(0.087)	(0.082)	(0.095)	(0.132)	(0.141)	(0.137)	(0.216)
N	997	997	997	997	997	997	997	997	997

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Earnings measured using the 2010 household instrument.

TABLE 8: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR SOCIO-EMOTIONAL FAMILY OUTCOMES

	Family Z-Score	Socio-emotional Family Components			
		Positive Self Evaluation	Optimism	Stress	Negative Self Evaluation
	(1)	(2)	(3)	(4)	(5)
ITT	0.053	0.249**	0.287***	0.170**	0.155*
	(0.039)	(0.093)	(0.078)	(0.071)	(0.086)
N	900	900	900	900	900
R^2	0.152	0.194	0.180	0.097	0.106

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Socio-emotional components are the first four factors resulting from exploratory factor analysis of all socio-emotional questions. The family z-score is calculated by averaging the z-score for the individual components. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 individual instrument.

TABLE 9: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR ALL FAMILIES BY STRATA

	Education		Learning Family Z-Score	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Absolute (5 % Trim)	Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ITT	0.667**	0.170**	0.360***	0.201**	0.081	0.223**	0.185***
	(0.250)	(0.072)	(0.127)	(0.088)	(0.094)	(0.085)	(0.055)
Four Highest Strata (=1) * ITT	-0.629*	-0.119	-0.294**	0.119	0.185	-0.049	-0.220***
	(0.336)	(0.093)	(0.145)	(0.138)	(0.119)	(0.106)	(0.079)
Test: ITT + Highest Strata * ITT = 0							
P-value	0.861	0.350	0.342	0.005	0.002	0.018	0.504
N	1,006	1,007	907	1,006	997	1,006	900
R^2	0.427	0.347	0.453	0.146	0.099	0.097	0.158

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. The strata are ordered from 1 to 7 with one being the poorest. Highest strata includes strata 4–7, hence coefficient on ITT indicates estimates on 3 poorest strata. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables in columns 1-2 and 4-6 measured using the 2010 household survey; variables in columns 3 and 7 measured using the 2010 individual survey.

TABLE 10: 2010 MATCHING ABSOLUTE IMPACTS FOR ALL FAMILIES

	Education		Learning Family Z-Score	Economic Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Rank	Absolute (5% Trim)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ATT	1.375***	0.379***	0.386***	0.185	0.066	0.106	0.103
	(0.347)	(0.088)	(0.123)	(0.137)	(0.096)	(0.106)	(0.082)
N	690	690	616	690	690	687	613

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Absolute effects compare early treatment to comparison group in 2010. Mean of grades attained in the comparison group is 5.3 in 2010. ATT biased adjusted estimator (Abadie and Imbens 2011) using five nearest neighbors. Z-scores are calculated using the mean and standard deviation of the late treatment group. Variables in columns 1-2 and 4-6 measured using the 2010 household survey; variables in columns 3 and 7 measured using the 2010 individual survey.

TABLE 11: 2005 ABSOLUTE IMPACTS ON EDUCATION

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1
	(1)	(2)	(3)	(4)
Treatment municipality * 2005 (δ3)	0.597***	0.124***	0.037***	0.091***
	(0.078)	(0.014)	(0.014)	(0.013)
N	18,399	18,399	18,421	18,403
R^2	0.107	0.088	0.037	0.061
Mean untreated municipalities in 2005	3.922	0.559	0.456	0.779

Notes: *** p<0.01, ** p<0.05, * p<0.10. 2005 absolute effects use national census data to compare rural areas of program municipalities to rural areas of comparison group municipalities. Heteroskedasticity-robust standard errors are given in parentheses. All variables measured using the 1995 and 2005 population censuses and include boys ages 9-12 in November of 1990 and 2000, respectively.

For Online Publication

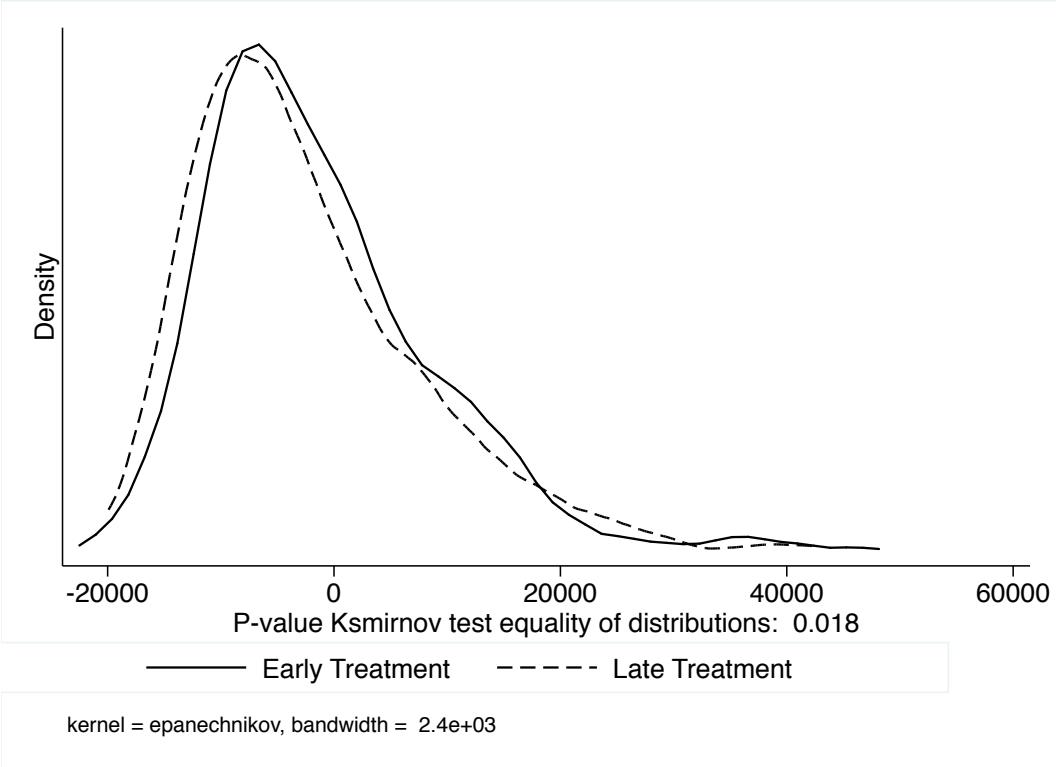
Online Appendices for "Are Conditional Cash Transfers Fulfilling Their Promise? Schooling, Learning, and Earnings After 10 Years"

By TANIA BARHAM, KAREN MACOURS, AND JOHN A. MALUCCIO

September 2018

APPENDIX A:

FIGURE A1: MONTHLY EARNINGS OFF-FARM, BOYS 9-12 IN 2000



Notes: Results for boys 9–12 in 2000. Earnings per month worked are demeaned using three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region and trimmed at the top five percent of values.

Source: Author's calculations.

TABLE A1: BASELINE BALANCE FOR 9–12 COHORT SAMPLE

	Early Treatment			Late Treatment			Diff. in Means		Mean/ SD
	Mean	SD	N	Mean	SD	N	Diff.	P-value	
Individual Characteristics									
Age at start of transfers in months	11.0	1.12	516	11.0	0.43	490	-0.02	0.67	-0.06
No grades attained (=1)	0.46	1.31	516	0.43	1.18	490	0.03	0.75	0.02
Highest grade attained	1.21	3.39	516	1.19	2.86	490	0.02	0.88	0.01
Worked last week (=1)	0.17	0.61	516	0.21	0.48	490	-0.04	0.16	-0.09
Mother not living in same household	0.08	0.28	516	0.07	0.33	490	0.01	0.43	0.03
Father not living in same household	0.22	0.66	516	0.18	0.40	490	0.04	0.16	0.11
Child of household head	0.86	0.41	516	0.88	0.35	490	-0.02	0.39	-0.06
Mother no grades attained (=1)	0.45	0.81	516	0.49	0.77	490	-0.04	0.32	-0.05
Mother 3 plus grades attained (=1)	0.37	0.85	516	0.32	0.75	490	0.04	0.33	0.06
Household Head Characteristics									
Age	44.8	14.6	516	44.4	13.2	490	0.40	0.58	0.03
No grades attained (=1)	0.53	0.80	516	0.50	0.55	490	0.04	0.33	0.06
3 plus grades attained (=1)	0.29	0.57	516	0.28	0.47	490	0.02	0.54	0.04
Worked last week (=1)	0.85	0.68	516	0.90	0.39	490	-0.05	0.06	-0.13
Household Characteristics									
Log predicted expenditures (pc)	7.71	0.73	516	7.74	0.68	490	-0.03	0.38	-0.04
Number of household members	8.26	4.57	516	8.22	5.20	490	0.04	0.91	0.01
Number of children aged 0-8	2.10	2.31	516	2.08	2.57	490	0.02	0.92	0.01
Number children aged 9-12	1.76	0.89	516	1.80	1.38	490	-0.04	0.53	-0.03
Log of size of landholdings	7.81	7.38	516	8.10	8.87	490	-0.29	0.57	-0.03
Family network size (individuals)	92.2	201.3	516	68.36	162	490	23.83	0.03	0.15
Own house (=1)	0.81	0.89	516	0.88	0.62	490	-0.08	0.11	-0.12
Some in household work in ag	0.82	0.78	516	0.85	0.86	490	-0.03	0.61	-0.03
Wealth index - housing characteristics	0.10	3.81	516	-0.02	2.90	490	0.13	0.57	0.04
Wealth index - productive assets	-0.13	1.53	516	0.14	2.08	490	-0.27	0.02	-0.13
Wealth index - other assets	-0.04	3.18	516	0.00	3.45	490	-0.03	0.78	-0.01
Number of rooms in house	1.59	1.65	516	1.58	1.34	490	0.01	0.89	0.01
Cement block walls (=1)	0.17	0.90	516	0.15	0.64	490	0.02	0.68	0.04
Zinc roof (=1)	0.54	1.60	516	0.49	1.39	490	0.05	0.62	0.04
Dirt floor (=1)	0.85	0.67	516	0.87	0.58	490	-0.02	0.64	-0.04
Latrine (=1)	0.65	1.25	516	0.60	1.14	490	0.04	0.57	0.04
Electric light (=1)	0.26	1.18	516	0.20	1.01	490	0.06	0.24	0.06
Radio (=1)	0.22	0.59	516	0.22	0.81	490	-0.01	0.86	-0.01
Work animals (=1)	0.11	0.35	516	0.18	0.73	490	-0.07	0.04	-0.09
Fumigation sprayer (=1)	0.31	0.81	516	0.38	1.15	490	-0.07	0.13	-0.06
Distance to nearest school (minutes)	24.9	80.5	516	23.9	66.8	490	1.09	0.84	0.02
Live in Tuma region (=1)	0.52	2.74	516	0.29	2.40	490	0.23	0.13	0.09
Live in Madriz region (=1)	0.19	2.07	516	0.18	2.19	490	0.01	0.89	0.00
Village Population	595	1962	516	335	143	490	259	0.01	0.18
Characteristics of Nearest School									
Highest grade school offers	4.78	3.34	512	4.85	6.12	484	-0.07	0.80	-0.01
Student-teacher ratio	36.7	28.2	507	36.8	40.7	419	-0.16	0.83	0.00
School under local governance	0.30	2.22	512	0.27	2.09	484	0.03	0.76	0.02

Notes: Standard errors and deviations are clustered at the locality level following the stratified randomization design. Means are weighted to account for sampling and attrition providing population estimates. The sample includes boys age 9–12 at program start with information in the 2010 household survey. Balance is similarly obtained for the sample of boys with information in the 2010 individual survey. The p-value for the difference in means includes controls for strata following the program design (difference in means do not include strata controls). The mean divided by the standard deviation uses the standard deviation of the late treatment group. The predicted per capita expenditures are from the program census data and uses the proxy means method developed by the government of Nicaragua for the purpose of household targeting (based on the 1998 Nicaraguan Living Standards Measurement Survey; Maluccio, 2009). The asset indices are constructed using principal component analysis for the assets list (appendix D). School characteristics are constructed from program administrative data collected for monitoring conditionalities. School under local governance refers to schools that participated in Nicaraguan's school autonomy reform, which provided schools and parents a certain level of autonomy over their own management and operations.

TABLE A2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR ALL FAMILIES, BY VARIOUS AGE GROUPS

	Education		Learning Family Z-Score	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Socio-Emotional Family Z-Score
	Grades Attained	Family Z-Score			Rank	Absolute (5 % Trim)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Differential Effects Age 7-8							
ITT	0.042	-0.008	-0.041	0.131	0.160	0.097	-0.013
	(0.245)	(0.079)	(0.080)	(0.091)	(0.096)	(0.091)	(0.045)
N	498	498	499	498	498	496	492
Panel B: Differential Effects Age 7-12							
ITT	0.207	0.060	0.107*	0.213***	0.172***	0.150**	0.035
	(0.178)	(0.051)	(0.061)	(0.057)	(0.059)	(0.059)	(0.033)
N	1,504	1,505	1,406	1,504	1,504	1,493	1,392

Notes: *** p<0.01, ** p<0.05, * p<0.10. Mean for late treatment for grades attained is 5.6 years. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables in columns 1-2 and 4-6 measured using the 2010 household survey; variables in columns 3 and 7 measured using the 2010 individual survey.

TABLE A3: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, WITH PROXY VARIABLES BY VARIOUS AGE GROUPS

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1	Ever Married =1	Worked Off-Farm =1 (last 12 months)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Differential Effect Age 9-12 - Population Weighted Including Proxy Reports						
ITT	0.353**	0.053**	0.028	0.049**	-0.055*	0.062**
	(0.157)	(0.021)	(0.018)	(0.018)	(0.032)	(0.025)
N	1,072	1,072	1,069	1,073	1,073	1,071
Mean Late Treatment	5.391	0.733	0.182	0.864	0.276	0.828
Panel B: Differential Effect Age 9-13 - Population Weighted Including Proxy Reports						
ITT	0.383**	0.059***	0.027	0.042***	-0.061**	0.049*
	(0.143)	(0.016)	(0.019)	(0.015)	(0.029)	(0.025)
N	1,293	1,293	1,288	1,294	1,296	1,291
Mean Late Treatment	5.352	0.724	0.172	0.857	0.314	0.825
Panel C: Difference in ITT effect between Age 13 and 9-12						
P-value	0.233	0.404	0.924	0.326	0.269	0.430
Panel D: Differential Effect Age 7-13 - Population Weighted Including Proxy Reports						
ITT	0.287**	0.046***	0.011	0.027*	-0.047**	0.034
	(0.140)	(0.016)	(0.028)	(0.014)	(0.021)	(0.022)
N	1,841	1,841	1,836	1,842	1,844	1,839
Mean Late Treatment	5.463	0.742	0.249	0.873	0.240	0.788

Notes: *** p<0.01, ** p<0.05, * p<0.10. Proxy reports included for all variables except enrolled because they were not collected. Regressions are weighted to account for sampling, but not attrition, providing population estimates. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling, but not attrition, providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household survey. For young men that were no longer member of their original household nor member of a split-off household, parental proxy reports obtained in the original household are included. [1] Because 13-year olds were not intensively tracked (appendix F) it is not possible to construct attrition weights similar to those done for the younger cohort.

TABLE A4: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR MARRIAGE STATUS AND FERTILITY

	First Had Sex by Age 15 (=1)	Ever Married =1	Any Children =1
	(1)	(2)	(3)
ITT	0.080**	-0.076**	-0.048
	(0.035)	(0.036)	(0.038)
N	875	1,007	875
R^2	0.111	0.081	0.096
Mean late treatment	0.269	0.290	0.225

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include boys ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 individual instrument, with the exception of marriage status.

APPENDIX B: ROBUSTNESS—ALTERNATIVE SPECIFICATIONS AND SAMPLES

In appendix Table B1, we examine the robustness of the 2010 differential results to a number of alternative specifications and samples. Results from the main specification are reproduced in panel A for comparison.

First, in panel B we exclude those children who were oversampled from the census for the 2010 survey, and only include people who were surveyed in the baseline survey, and then again in the 2010. The number of observations is reduced by a little more than a half, and hence standard errors increase, but all significant results remain significant and the point estimates are broadly similar.

Second, we limit the sample to 11-year olds at the start of the program in Panel C, the group we oversampled because of their large potential differential exposure. We expect program effects to be larger for this group than for the overall sample, but possibly less precise as the sample size is substantially smaller (353 versus 1,007). Results indeed show higher point estimates for educational gains (0.7 grades attained) and learning (0.3 SD). The point estimates on earnings are similar to the main results (approximately 0.2 SD), but less precisely estimated.

Third, we estimate effects controlling only for strata to follow the stratified randomized design of the program. While this also increases standard errors, as expected, the results in panel D are broadly consistent with the main results.

Fourth, in panel E we include additional baseline controls—estimated per capita expenditures, number of children 9–12 in the household, distance to school, the productive assets component of the wealth index, an indicator for whether the child worked at baseline, and family network size. These are important correlates of the outcomes, and the final three were not balanced at baseline. Results are similar to the main results.

Fifth, we explore the sensitivity of the analyses to the attrition correction weights (see appendix F for more details). The main results in the paper are weighted to account for attrition, by multiplying the attrition corrected weights and sampling weights. To examine the results without the attrition weights, we present results using only the sampling weights in panel F. Tables B2.1-B2.4 repeat the more detailed analysis of the variables in each family (education, learning, labor market participation, and earnings) without attrition weights. Alternatively, we also calculate a different set of attrition weights using individuals found during both the regular and the intensive tracking phases, and show inverse probability weighted regressions with these alternative attrition weights multiplied by the sampling weights in Panel G. All findings are robust to only using sampling weights (Table B1.1, Panel F and Tables B2.1-B2.4). The labor market, learning and earning results are also robust to using alternative attrition weights (Panel F), but increased standard errors turn the education result insignificant in this specification.

Sixth, in panel H we re-estimate effects for the variables derived from the household instrument for the sample of boys for whom there is information from the individual instrument, and use the weights designed to account for attrition and sampling in the individual survey. Results are qualitatively similar even if the earnings estimates are slightly lower and no longer significant.

Finally, panel I presents Lee (2009) bounds, where the bounds assume the tracked sample is either entirely negatively or entirely positively selected (the monotonicity assumption). The Lee bounds are estimated based on ITT estimations without any covariates (including without controls for strata). Given limited sample size, we are unable to trim the samples using covariates to obtain tighter bounds. Omitting covariates and strata controls increases point estimates for the

grades attained and learning results, with both upper and lower bounds significantly different than zero. Upper and lower bound for earnings results are also significant. On the other hand, omitting covariates and strata controls reduces the size of the point estimates for the labor market participation family, and the lower bound becomes insignificant. For the education family, the lower bound also is insignificant.

Appendix Table B1.2 further examines the robustness of the inference. Following Anderson (2008), we adjust the p-values for the possibility of multiple hypotheses testing using the familywise error rate in panel A. With the exception of the socio-emotional outcomes, all differential estimates are significant at 5 percent or below.

Panel B shows exact p-values obtained through randomization inference, by calculating all possible realizations of the test statistic and rejecting if the observed realization in the experiment itself is extreme enough, using Young (2017). This is potentially important given the relatively small number of randomized clusters (42) in the experiment. Results confirm we can reject that the Fisher exact null-hypotheses of no differential treatment effects for all families except the socio-emotional outcomes.

TABLE B1.1: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, ALTERNATIVE SPECIFICATIONS

	Education		Learning Family Z-Score	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Rank	Absolute (5% Trim)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Age 9–12 Main results							
ITT	0.288*	0.098**	0.183**	0.272***	0.194***	0.192***	0.053
	(0.167)	(0.043)	(0.070)	(0.075)	(0.057)	(0.067)	(0.039)
N	1,006	1,007	907	1,006	1,006	997	900
Panel B: Age 9-12 Excluding Over-Sample Children							
ITT	0.375*	0.124**	0.172**	0.266**	0.224***	0.214**	0.027
	(0.197)	(0.061)	(0.069)	(0.118)	(0.081)	(0.104)	(0.056)
N	496	497	444	497	497	491	441
Panel C: Age 11 Only							
ITT	0.681*	0.177*	0.283***	0.283**	0.183	0.223	-0.046
	(0.362)	(0.099)	(0.079)	(0.127)	(0.129)	(0.135)	(0.070)
N	352	353	321	353	353	349	318
Panel D: Age 9–12 Strata Controls Only							
ITT	0.361	0.109	0.326**	0.165	0.133**	0.165**	0.128*
	(0.324)	(0.078)	(0.141)	(0.099)	(0.059)	(0.063)	(0.070)
N	1,006	1,007	907	1,006	1,006	997	900
Panel E: Age 9–12 Extended Controls							
ITT	0.289*	0.105**	0.165**	0.240***	0.179***	0.182***	0.045
	(0.167)	(0.044)	(0.067)	(0.069)	(0.057)	(0.064)	(0.039)
N	1,006	1,007	907	1,006	1,006	997	900
Panel F: Age 9–12 Sampling Weights (No attrition correction)							
ITT	0.351**	0.105**	0.175***	0.266***	0.195***	0.188***	0.033
	(0.153)	(0.040)	(0.063)	(0.070)	(0.057)	(0.069)	(0.041)
N	1,006	1,007	907	1,006	1,006	997	900
Panel G: Age 9-12 Attrition Correction using Observations From Intensive & Regular tracking							
ITT	0.296	0.089	0.164***	0.282***	0.195***	0.181**	0.0450
	(0.188)	(0.054)	(0.062)	(0.068)	(0.073)	(0.08)	(0.047)
N	1,006	1,007	907	1,006	1,006	997	900
Panel H: Restrict Household Survey Variables to Individual survey Sample and Weight							
ITT	0.351*	0.138***		0.245***	0.100	0.121	
	(0.176)	(0.048)		(0.080)	(0.071)	(0.077)	-
N	906	907		907	907	898	
Panel I: Lee bounds							
Lower Bound	0.456*	0.018	0.237***	-0.084	0.152**	0.149**	0.043
	(0.237)	(0.060)	(0.077)	(0.070)	(0.072)	(0.072)	(0.053)

Upper Bound	0.707***	0.180***	0.310***	0.184**	0.200**	0.239***	0.090
	(0.249)	(0.062)	(0.090)	(0.082)	(0.081)	(0.091)	(0.059)

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and in parentheses. Regressions include boys ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates, unless otherwise stated. Differential ITT results compare early to late treatment groups in 2010. The family z-score averages the z-scores for the individual components and is calculated if any component is available. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables in columns 1-2 and 4-6 are measured using the 2010 household survey; variables in column 3 and 7 are measured using the 2010 individual survey.

TABLE B1.2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, ROBUSTNESS INFERENCE

	Education Family Z-Score	Learning Family Z-Score	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
				Rank	Absolute (5% Trim)	
	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Multiple Hypothesis Testing – Familywise Error Rate Adjusted p-values						
p-value	0.034	0.020	0.005	0.005	0.013	0.186
Panel B: Randomization Inference						
Exact p-value	0.036	0.030	0.002	0.001	0.008	0.256

Notes: Regressions include boys ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score averages the z-scores for the individual components and is calculated if any component is available. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables in columns 1-2 and 4-6 are measured using the 2010 household survey; variables in column 3 and 7 are measured using the 2010 individual survey. Panel A adjusts the p-values for multiple hypothesis testing using the familywise error rate following Anderson (2008) based on the variables included in the table. Panel B shows Fisher exact p-values obtained through randomization inference using Young (2017)'s randomization-t.

TABLE B2.1: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EDUCATION FAMILY AND LITERACY, SAMPLING WEIGHTS

	Education Family Z-Score	Education Family Components			Read and Write =1
		Grades Attained	Completed Grade 4 =1	Enrolled =1	
	(1)	(2)	(3)	(4)	(5)
ITT	0.105**	0.351**	0.048**	0.033	0.053***
	(0.040)	(0.153)	(0.021)	(0.020)	(0.019)
N	1,007	1,006	1,006	1,005	1,007
R^2	0.321	0.420	0.263	0.089	0.165
Mean late treatment		5.452	0.743	0.185	0.865

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score averages the z-scores for the individual components and is calculated if any component is available. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables measured using the 2010 household survey.

TABLE B2.2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LEARNING AND COGNITION FAMILIES (Z-SCORES), SAMPLING WEIGHTS

	Learning			Mixed Cognition and Learning	Cognition (Raven)
	Math and Spanish	Math	Spanish		
	(1)	(2)	(3)	(4)	(5)
ITT	0.175***	0.135**	0.205***	0.109*	0.033
	(0.063)	(0.064)	(0.069)	(0.061)	(0.077)
N	907	905	907	906	906
R^2	0.375	0.337	0.354	0.288	0.198

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables measured using the 2010 individual survey.

TABLE B2.3: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LABOR MARKET PARTICIPATION AND MIGRATION, SAMPLING WEIGHTS

	Labor Market Participation Family Z-Score	Labor Market Participation Family Components				Permanent Migration Out of Municipality =1
		Worked Off-Farm =1 (last 12 months)	Migrated for Work =1 (last 12 months)	Ever Had a Salaried Non-Agricultural Job =1	Ever Worked in Urban Area =1	
	(1)	(2)	(3)	(4)	(5)	(6)
ITT	0.266***	0.060***	0.086***	0.083**	0.066**	-0.026
	(0.070)	(0.022)	(0.027)	(0.034)	(0.032)	(0.021)
N	1,006	1,006	1,006	998	998	1,007
R^2	0.140	0.067	0.165	0.125	0.118	0.037
Mean late treatment		0.828	0.309	0.221	0.118	0.115

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables measured using the 2010 household survey.

TABLE B2.4: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EARNINGS FAMILY AND COMPONENTS, SAMPLING WEIGHTS

	Family Z-Score	Earnings Family Components (C$)			
		Earnings Per Month Worked (last 12 months)	Annual Earning (last 12 months)	Maximum Earnings (last 12 months)	Maximum Salary Ever
	(1)	(2)	(3)	(4)	(5)
Panel A: Rank of Earnings					
ITT	0.195***	43.002**	24.689	47.126**	45.817**
	(0.057)	(19.421)	(18.515)	(19.653)	(18.666)
N	1,006	1,006	1,006	1,006	998
R^2	0.090	0.075	0.083	0.081	0.107
Mean late treatment		493.3	492	494.5	482.6
Panel B: Earnings — Five Percent Trim					
ITT	0.188***	203.325***	555.966	196.323**	157.602**
	(0.069)	(68.572)	(630.163)	(72.834)	(72.671)
N	997	956	956	956	955
R^2	0.093	0.076	0.080	0.092	0.079
Mean late treatment		1437	7996	1626	219

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Earnings include wage work off the family farm. Earnings in panel A are trimmed at the top five percent of values. Earnings are in Nicaragua Cordobas (C$) and the exchange rate is approximately 20. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables measured using the 2010 household survey.

APPENDIX C: TARGETING AND DESIGN OF THE NICARAGUAN CCT PROGRAM [33]

The Nicaraguan CCT, *Red de Protección Social* (*RPS*), was designed to address both current and future poverty through cash transfers targeted to poor and extremely poor households in rural Nicaragua. Its specific stated objectives included: 1) supplementing household income for up to three years to increase expenditures on food; 2) reducing dropout rates during the first four years of primary school; and 3) increasing the nutritional status and healthcare of children under five. Implemented by the Government of Nicaragua with technical assistance and financial support from the Inter-American Development Bank (IDB), the program began in 2000 and had two budgetary phases over six years. The first phase lasted three years with a budget of $11 million. In late 2002, based in part on the positive findings of various program evaluations, the Government of Nicaragua and the IDB agreed to a continuation and expansion for a second phase until 2006, with an additional budget of $22 million. A randomized evaluation was included in the initial program design starting in 2000 carried out by the International Food Policy Research Institute (IFPRI).

Program Targeting—The CCT first targeted the rural areas in six municipalities in central and northern Nicaragua, from three regions, on the basis of poverty as well as on local capacity to implement the program. The focus on rural areas reflected the distribution of poverty in Nicaragua—of the 48 percent of Nicaraguans identified as poor in 1998, 75 percent resided in rural areas (World Bank 2001). While the six targeted municipalities were not the poorest in the country, the proportion of impoverished people living in these areas was still well above the national average (World Bank 2003). At the same time, the selected municipalities had relatively good communication and access (for example, less than one day's drive to Managua, where the program's central administrative office was located), relatively strong institutional capacity and local coordination, and adequate access to primary schools.

In the next stage of (pre-program) targeting, a marginality index was constructed for all 59 of the rural census *comarcas* (hereafter localities)[34] within the six selected municipalities. The index was the weighted average of a set of locality-level indicators (including average family size, lack of access to potable water and latrines and illiteracy rates); localities with higher marginality index scores were considered more impoverished.[35] The 42 localities with the highest scores were selected for inclusion in the randomized program evaluation (divided into the early and late treatment groups). Finally, while the initial program design called for all households in these 42 targeted localities to be eligible for the CCT, prior to the start of the program the government excluded approximately three percent of them determined to have substantial resources, in particular those who owned a vehicle or had large landholdings. These households are excluded from the analyses in the paper.

---

[33] This appendix draws on IFPRI (2005), Maluccio and Flores (2005) and Maluccio (2009).

[34] Census *comarcas* are administrative areas within municipalities based on the 1995 Nicaraguan national census that included as many as 10 small communities for a total of approximately 250 households.

[35] More specifically, the marginality index for each locality included average family size (10 percent), percent without piped water in the home or yard (50 percent), percent without a latrine (10 percent) and percent of persons over age five who were illiterate (30 percent)—all calculated from the 1995 Nicaraguan *National Population and Housing Census*.

While not statistically representative of rural Nicaragua as a whole, the 42 localities comprising the randomized evaluation area were nevertheless similar to other rural areas in the targeted regions and elsewhere in the country. For example, three-quarters of the approximately 1,000 rural localities in the country had marginality index scores in the same range as the program areas. By way of comparison, poverty rates in the targeted localities were about 10 percentage points higher than rural national rates: 80 versus 69 percent poor and 42 versus 29 percent extremely poor.

During its operation, the CCT also expanded to the remaining 59 – 42 = 17 rural localities not initially targeted. In those 17 localities, which were less poor according to the marginality index, the CCT began in late 2001 and was offered to 80 percent of the population based on a household-level proxy means targeting model. Consequently during the period 2000-05, the three-year program had been implemented (to modestly different degrees and at different times) in all (59) rural localities of the six municipalities, and more than 90 percent of the population had been eligible.

Program Components and Conditionalities—The CCT had two core components: 1) education; and 2) food security, nutrition, and health. Corresponding to these, transfers were conditional on household education and health behaviors, with conditionalities monitored by teachers and specially contracted healthcare providers. Conditionalities and benefits were first explained to eligible families in the early treatment group during registration assemblies in September and October 2000 and transfers began in November 2000. Only the designated household representative (referred to in Spanish as the *titular*) could receive the transfers and, where possible, the CCT appointed the mother or other female caregiver to this role. As a result, more than 95 percent of the household representatives were women. The CCT also worked with local volunteer coordinators (beneficiary women chosen by the community and referred to in Spanish as the *promotora*) to help implement the program. The coordinators organized and informed their group of household representatives regarding upcoming program activities, upcoming transfer payments, and failure to fulfill the conditions. Conditions were monitored for compliance, and if not reported by the school or healthcare provider as having been met, relevant transfers were withheld by the program. The program also had a strong social marketing message that the money was intended to be used for food and education investments and beneficiaries were required to sign a short agreement to that effect although explicit expenditures were not actually monitored.

Education component. Each eligible household received a bimonthly (every two months) cash transfer known as the school attendance transfer, contingent on enrollment and regular school attendance of children aged 7–13 years who had not yet completed the fourth grade of primary school. For each eligible child, the household also received an annual cash transfer at the start of the school year (which begins in January) intended for school supplies (including uniforms and shoes) known as the school supplies transfer, which was contingent on enrollment. Unlike the school attendance transfer, a fixed amount per household regardless of the number of eligible children, the school supplies transfer was a per child transfer.

To provide incentives to the teachers and increase resources available to the schools, there was also a small cash transfer, known as the teacher transfer. In rural Nicaragua, school parents associations often request small monthly contributions from parents to support the teacher and the school; the teacher transfer was, in part, intended to substitute for this type of fee. This transfer was per child as well, and delivered directly to the household, which was then required

to pass the funds along to the teacher. The instruction was that the teacher keep half with the other half earmarked for the school. Although the delivery of the funds by the household to the teacher was a program condition that was monitored, the ultimate use of the funds was not. The teacher transfer was continued in areas even after household transfers had stopped until the end of the program. Teachers and schools completed specially designed scan forms that were regularly submitted to the central office to verify conditions, assess compliance, and determine transfers.[36]

While there was no explicit supply-side intervention for education such as a school building program (having targeted the program to areas with adequate schooling infrastructure), the centrally administered CCT had a multisectoral approach promoting inter-institutional cooperation through specially formed committees at the national, municipal, and local levels. This coordination proved useful in some areas for directing ad hoc supply-side responses to increased demand, including the placement of additional government teachers.

Food security, nutrition, and health component. Each eligible household received a bimonthly cash transfer known as the nutrition and health transfer that was a fixed amount per household, regardless of household size and regardless of whether a household had children subject to the associated conditionalities. The transfer was contingent upon the household representative attending bimonthly health education workshops and bringing children under age five for scheduled preventive healthcare appointments. The workshops were held within the communities and covered household sanitation and hygiene, nutrition, reproductive health, and related topics. The required preventive healthcare appointments were scheduled monthly for children under age two and bimonthly for those age two to five. Health services at the scheduled visits included growth monitoring, vaccination, and provision of iron supplementation drops and anti-parasite medicine.

The program supplemented the supply of specialized healthcare services in the areas to ensure that increased demand could be met without reducing quality. Specifically, the CCT contracted and trained private healthcare providers to deliver the program-related services free of charge (Regalia and Castro 2007), and beneficiaries were required to use those providers for fulfillment of the conditions. Providers visited program areas on scheduled dates and delivered services in existing health facilities, community centers or private homes. They completed specially designed scan forms recording the services delivered that were regularly submitted to the central office to verify conditions, assess compliance, and determine transfers.

In 2003, as the early treatment group was phasing out and the late treatment group phasing in, a number of additional services and corresponding conditions were added. Chief among them were vaccination for school-age children, family planning services for women of childbearing age, prenatal care consultations, and an additional set of health education workshops for adolescents. All adolescents were required to attend the additional workshops (segregated by age groups) and they focused on healthy living and reproductive health, including contraception. At the same time, modern contraceptive methods were made available to beneficiaries through the healthcare providers. These additional services were designed to be implemented after 2003 in both the early and the late treatment groups, but attendance was a conditionality for transfer payments only for the late treatment group as transfers to the early group were ending in 2003. Thus all girls 9-12 in 2000 were eligible for these sessions in the late treatment group. In

[36] Although not an explicit component, the CCT administration did work with localities on an ad hoc basis to alleviate bottlenecks in assignment of new teachers and in the second phase offered some limited teacher training.

practice, the services were only partially implemented in the early treatment group because there were fewer synergies with other ongoing program components. Consequently, attendance of adolescents at the health education workshops was lower in the early compared to late treatment groups.

Transfer Sizes—The initial annual transfer amounts in U.S. dollars (using the September 2000 average exchange rate of C$ 12.85 Nicaraguan Córdobas to US$ 1.00) were as follows: the nutrition and health transfer was $224 a year; the school attendance transfer $112; the school supply transfer $21; and the teacher transfer $5. On its own, the nutrition and health transfer represented about 13 percent of total annual household expenditures in beneficiary households before the program. A household with one child benefiting from the education component would have received additional transfers of about 8 percent, yielding an average total potential transfer of 21 percent of total annual household expenditures. On average, transfers made were 18 percent of pre-program expenditures in the first phase. The nominal value of the transfers remained constant, with the consequence that the real value of the transfers declined by about eight percent due to inflation during the first budgetary phase. The size of the transfers was reduced for the late treatment group. For that group, the nutrition and health transfer started at $168 for the first year of program participation and then declined to $145 and $126 in the second and third years. The school attendance transfer also declined slightly, to $90 per year, but the school supplies and teacher transfers increased to $25 and $8 per eligible child. These figures represent potential transfer amounts, i.e., the transfer amount received upon complying fully with all associated conditions.

To enforce compliance with program requirements, beneficiaries did not receive the nutrition and health, or separately education, transfers, in a given transfer period when they failed to carry out all of the relevant conditions described above. Compliance was measured via the reporting from schools and the private healthcare providers. Repeated violation, including two consecutive periods of non-compliance, led to households losing their overall eligibility.

APPENDIX D: WEALTH INDEX

The baseline program census data contain a number of variables to proxy for household wealth, including characteristics of the housing structure and household assets. Following Filmer and Pritchett (2001), we aggregate these characteristics using principal components analysis. The principal components are estimated using the baseline target sample of all 9–12 year olds boys regardless of whether they were interviewed in 2010. We retain the first three principal components as they each have an eigenvalue of more than one; jointly, they account for 53 percent of the variation of the nine underlying variables included. The first principal component mostly captures characteristics of the house, the second productive assets (ownership of work animals and a fumigation sprayer), and the third specific household amenities (roof materials and latrines).

TABLE D1. PRINCIPAL COMPONENT SCORING COEFFICIENTS

Variable	PC 1	PC 2	PC 3
Household Characteristics			
Work animals (=1)	0.13	0.62	-0.03
Fumigation sprayer (=1)	0.12	0.59	0.36
Number of rooms in the house	0.35	0.27	0.01
Radio (=1)	0.39	0.07	-0.21
Cement block walls (=1)	0.44	-0.10	0.11
Zinc roof (=1)	0.21	-0.16	0.69
Dirt floor (=1)	-0.44	0.26	-0.18
Latrine (=1)	0.28	0.14	-0.49
Electricity light (=1)	0.43	-0.27	-0.23

Notes: PC refers to principal component

APPENDIX E: LEARNING, COGNITIVE, AND SOCIO-EMOTIONAL OUTCOMES

All standardized tests included in the 2010 individual survey instrument were piloted extensively and minor adjustments made for the local context as necessary, such as rephrasing questions for maximum understanding. Similar tests have been applied in other populations in Latin America, including in the evaluations of CCT programs in Ecuador and Mexico, and a different CCT program in Nicaragua (Behrman, Parker, and Todd 2009a; Fernald, Gertler, and Neufeld 2009; Paxson and Schady 2010; Macours, Schady, and Vakis 2012).

Tests were conducted in the young adult respondents' homes by specially trained female test administrators. Therefore, the results were obtained independent of whether the respondent was in school, avoiding potential selection concerns.

Test administrators were selected for their background (trained as psychologists, social workers, or similar fields) and for their ability to quickly establish a strong rapport with children and young adults. They were trained to motivate the respondents to participate in the tests, keeping final non-response to a minimum. Tests were administered inside the home (or in the compound) and the privacy of the test-taker and the confidentiality of the results were assured throughout the process. During the test administrators' training, emphasis was placed both on gaining the confidence of the respondents before starting the tests and on the standardized application of each of the tests. The quality and standardized application of the tests was monitored closely in the field, and given the long survey period, several re-standardization trainings were carried out.

Data collection and test administration was organized in such a way that the test administrators would maintain a balance between the number of children visited in early and late treatment localities. Visits to early and late treatment localities were also balanced over time to avoid possible seasonal differences in measurement between the experimental groups. Consistent with these field protocols, the main results are robust to controls for the identity of the test administrator (not shown).

All tests are scored using standard practice. For the digit span, we use the combined score of the forward and backward components of the test. When estimating the differential ITT on each component separately, estimates of the forward digit span (measuring math memory) are in line with the TVIP and significant, while the backward digit span (often considered a measure of executive functioning) is not significant at all and similar to the Raven.

Exploratory Factor Analysis for Socio-Emotional Outcomes—Two standardized instruments to measure socio-emotional outcomes were applied in the individual instrument. The first was the Strength and Difficulties test (SDQ), a self-reported behavioral screening test consisting of 25 questions aimed at measuring a set of positive and negative behaviors. In addition, we implemented the Center for Epidemiologic Studies Depression Scale or CESD (Radloff 1977), a commonly used mental health scale, developed as a screening test for depression and depressive disorder and consisting of 20 questions asking for the frequency of both positive and negative self-perceptions. Both tests are available in Spanish.

We first analyze the internal consistency for the sample of boys studied (9–12 year olds at baseline) through exploratory factor analysis. The overall Cronbach's alpha of the 25 items of the SDQ together (0.70) indicates that the scale as a whole is internally consistent. But the alphas are much lower when considering the five usual subdomains (emotional symptoms, conduct problems, hyperactivity, peer relationships, and pro-social behavior), and vary from 0.26 to 0.51,

hence much lower than the usual threshold for statistical validity. Exploratory factor analysis on the 25 items suggests there are only two factors that can be meaningfully retained (i.e., two factors have eigenvalues above one and the scree plot leads to a similar conclusion). Moreover, when imposing the 5-factor structure, the items do not group along the original five subscales, with the first factor having high factor loads on items from three of the five subscales. When we consider the CESD, the Cronbach's alpha for internal consistency of the 20 items is high (0.83) but the factor analysis only points to one or two factors, and does not allow further differentiation.

As the data suggest that we should not pool questions together based on regular subcategories of the SDQ, we construct new indices capturing the relevant latent traits, based on all items in the SDQ and CESD. We pool together all questions from the SDQ and the CESD scales and identify the latent socio-emotional traits in our sample, following, among others, Cunha, Heckman, and Schennach (2010) and Attanasio et al. (2015). Based on both the eigenvalue and the scree plot, and using an oblique quartimin rotation to allow the different factors to be correlated with one another, we retain four factors, two factors with high loads on items from the SDQ scale, and two factors with high loads on items from the CESD scale. Notably, questions referring to positive, respectively negative, attitudes or behavior are pooled in each of the scales. Hence considering the factor loadings of the different items points to a plausible interpretation of these factors as capturing stress, positive self-evaluation, negative self-evaluation, and optimism.[37] Results are shown in appendix Table E1.

[37] All results are qualitatively similar with or without inclusion of the non-experimental comparison group.

TABLE E1. FACTOR LOADINGS OF SOCIO-EMOTIONAL QUESTIONS

	Factor 1 Stress	Factor 2 Positive Self-Evaluation	Factor 3 Negative Self-Evaluation	Factor 4 Optimism
CESD				
During the last 7 days, how many days …				
… were you bothered by things that usually don't bother you?	0.45	0.08	-0.10	0.12
… did you not feel like eating? (your appetite was poor)	0.53	-0.06	0.09	-0.07
… did you feel that you could not shake off the blues even with help from your family and friends?	0.66	-0.06	0.06	-0.01
… did you feel that you were just as good as other people?	0.60	0.05	0.00	-0.05
… did you have trouble keeping your mind on what you were doing?	0.56	0.01	-0.03	0.14
… did you feel depressed?	0.58	0.04	-0.06	0.08
… did you feel that everything you did was an effort?	0.38	-0.02	-0.04	0.23
… were you hopeful about the future?	0.16	-0.07	0.01	0.54
… did you think your life had been a failure?	0.56	0.08	0.01	-0.07
… did you feel fearful?	0.53	-0.03	-0.02	0.00
… was your sleep restless?	0.48	-0.02	0.05	0.09
… were you happy?	-0.11	0.00	0.01	0.47
… did you talk less than usual?	0.39	-0.04	-0.02	0.10
… did you feel lonely?	0.56	-0.01	-0.02	-0.06
… people were unfriendly?	0.54	0.10	-0.06	0.06
… did you enjoy life?	-0.29	0.01	0.06	0.41
… did you have crying spells?	0.45	-0.01	0.03	-0.13
… did you feel sad?	0.62	-0.07	-0.04	-0.16
… did you feel that people disliked you?	0.48	0.05	-0.14	0.06
... could you not get 'going'?	0.49	-0.03	-0.04	0.01
… did you feel you were moving ahead in life?	-0.05	-0.03	0.02	0.61
… where you thinking abou the way to move ahead in life?	0.04	-0.09	0.01	0.59
SDQ				
I try to be nice to other people. I care about other people's feelings	0.01	0.19	0.10	-0.21
I am restless, I cannot stay still for long	0.02	0.15	0.07	-0.15
I get a lot of headaches, stomach-aches or sickness	-0.04	0.14	0.25	0.11
I usually share with others, for example food, pencils/	0.01	0.23	0.17	-0.08
I get very angry and often lose my temper	-0.04	0.00	0.43	0.04
I would rather be alone than with other people	-0.05	0.16	0.18	0.07
I usually do as I am told	-0.01	0.41	-0.09	0.01
I worry a lot*	-0.02	0.39	0.08	-0.03
I am helpful if someone is hurt, upset or feeling ill	0.00	0.48	0.01	-0.02
I am constantly fidgeting or squirming*	0.13	0.32	0.13	-0.13
I have one good friend or more	0.06	0.34	0.02	0.04
I fight a lot. I can make other people do what I want	-0.02	-0.15	0.35	-0.05

I am often unhappy, depressed or tearful	-0.14	0.13	0.41	0.01
Other people my age generally like me	0.05	0.42	-0.03	0.00
I am easily distracted, I find it difficult to concentrate	0.05	0.30	0.16	0.11
I am nervous in new situations. I easily lose confidence	-0.05	0.11	0.35	0.17
I am kind to younger children	0.02	0.36	0.08	-0.11
I am often accused of lying or cheating	0.04	-0.07	0.47	-0.06
Other young people pick on me or bully me	-0.07	-0.03	0.45	0.06
I often offer to help others (parents, teachers, children)	-0.07	0.53	-0.02	0.00
I think before I do things	-0.03	0.46	-0.01	-0.07
I take things that are not mine from home, school or elsewhere	0.04	-0.16	0.32	-0.03
I get along better with adults than with people my own age	0.03	0.39	0.03	-0.01
I have many fears, I am easily scared	0.00	0.19	0.38	0.06
I finish the work I'm doing. My attention is good	0.05	0.46	-0.06	-0.13

Notes: * denotes items that are meant to capture negative traits (difficulties) in English but in the Spanish translation may have been interpreted as positive by the respondents.

APPENDIX F: TRACKING AND ATTRITION CORRECTION

In the 2010 survey, we placed special emphasis on tracking all migrants and other difficult-to-interview individuals. In the first phase of the survey, lasting about 6 months, we interviewed individuals and households located in or nearby their original localities. We refer to this period as the "regular tracking" phase. This was followed by an "intensive tracking" phase, lasting approximately 1.5 years, during which we made exhaustive efforts to find all individuals not found during regular tracking, through repeat visits to original locations and tracking to any location in Nicaragua or Costa Rica. For individuals who could not be located, however, some information on selected individual variables is available, having been collected through proxy reports when interviewing the original household.

We hence distinguish between three sets of outcomes based on their source—from the household-level instrument, from the individual-level instrument, or by proxy from the household-level instrument. Outcomes on individuals collected in the household instrument include educational attainment, all labor market and earnings outcomes, and marital status, self-reported by the individual or—in cases when he was resident but temporarily absent—another informed household member. Outcomes collected in the individual-level instrument include achievement and cognitive tests, socio-emotional outcomes, fertility, and reproductive health behaviors. And outcomes on which proxy information was collected in the original household include highest grade attained, enrollment, literacy, marital status, and labor market status.

Attrition is highest for the outcomes that required direct in-person interactions with the respondents themselves for the individual-level instrument. For the main sample of 9-12 cohort of boys examined in this paper, 19 percent could not be tracked for the individual instrument, 10 percent for the household instrument, and for 4 percent we are also missing a proxy report.[38]

Given the relatively small sample sizes, we intensively tracked all migrants rather than a random subset. Tracking rates in the intensive phase are comparable to intensive tracking rates obtained in other studies for random subsamples, resulting in a final tracking rate of 90 percent for the household instrument (and hence for the main labor market and earnings outcomes in the paper), comparable to other long-term follow-ups of RCTs with both regular and intensive tracking phases.[39] Attrition rates for our main outcomes are lower than in related studies analyzing the long-term experimental impacts of CCTs.[40]

---

[38] A small number of individuals (15) in the target age group were deceased by 2010. They are not used to predict the probability of attrition as selection for them is most likely driven by other factors. Including the deceased individuals, final attrition rates are 12 percent for the household and 20 percent for the individual instruments.

[39] It is, for instance, comparable to: Blattman, Fiala, and Martinez (2014) who report an 82 percent effective tracking rate for a 4-year follow-up survey of young adults in Uganda; the 10-year follow-up of the Kenya Longitudinal Panel Survey with an effective tracking rate of 84 percent (Baird et al., 2016); and the 88 percent effective tracking rate for children after 5-7 years in the Moving to Opportunity evaluation in the United States (Orr, 2003). Our tracking efforts, however, were less successful than the 8-year follow-up of a scholarship program in Ghana (Duflo, Dupas, and Kremer, 2017), who like us use intensive tracking for the entire sample of those not found during regular tracking, but where special protocols to track respondents were incorporated into the RCT design from baseline including maintaining regular contact with respondents throughout the duration of the study and resulting in a 98 percent tracking rate.

[40] Behrman, Parker, and Todd (2011) and Adhvaryu et al. (2018) use the 6-year follow up of Mexico's PROGRESA evaluation sample, with an attrition rate of 40 percent for individual-level information on comparable age groups. The 10-year PROGRESA follow-up (used for instance in Kugler and Rojas, 2018) has more than 60 percent attrition. Attrition rates in Baird, Mcintosh, and Ozler (2016) are 13-16 percent after five years for young women,

Table F1 shows that attrition rates are well balanced between the early and late treatment groups with the coefficients on the ITT indicator on the probability of having been found, i.e., interviewed, smaller than |1.5| percentage points for both the household and individual survey instruments. Table F2 confirms that this resulted in a sample that maintained balance on baseline characteristics, with 4 (respectively 6) out of the 42 baseline variables examined significantly different at the 10 percent level for the *final samples* with the individual (columns 5 and 6) and household (columns 9 and 10) instruments. These are the same differences observed for the full sample (columns 1 and 2), and hence do not appear to be driven by selective attrition.

To further assess the possibility of selective attrition, for each baseline variable we estimate $Prob(found) = F(X, T, XT)$. As seen in Table F3, only a few of the estimated coefficients on T are significant, but the coefficients on X clearly show that those who attrited are different than those found along a number of dimensions. The coefficients on the interaction effect further suggest that the observable characteristics of individuals (in particular baseline household demographics and village characteristics) of those who attrited are somewhat different between the early and late treatment groups.

In our preferred specifications, we explicitly account for such selection using inverse probability weights constructed as described below. We also present a number of complementary estimations to examine the sensitivity of the findings to different assumptions about attrition.

Attrition selection correction with inverse probability weights[41]

Because several baseline characteristics are correlated with the probability of having been found, at least some of the potential attrition selection is likely related to observables. Appendix Table F2 shows that prior to the intensive tracking phase, there were more baseline characteristics that were not balanced (columns 3 and 7) than would be expected by chance, but that baseline balance was restored through the intensive tracking (columns 5 and 9).[42] This holds for both household- and individual-level data. Intensive tracking hence proved important for internal validity. Those found during the intensive tracking phase differ from those found during the regular tracking phase in terms of observed baseline characteristics. Moreover, individuals that were difficult to find, but were ultimately found, are more similar on baseline observable characteristics to those that were ultimately not found. Among other differences, administrative data indicate that those that were more difficult to find had lower compliance rates with the CCT, suggesting that attrition correction may be important to account for heterogeneity in the ITT effects.

We therefore use a modified version of the more standard inverse probability weighting adjustment leveraging the information obtained during intensive tracking phase and putting more weight on individuals who were more difficult to find. The key assumption underlying the estimation strategy is that the probability of being found during the intensive tracking phase is explained by observable characteristics. Overweighting individuals whose observed characteristics predict they were difficult to find corrects for the sample selection. To determine

---

and the 10-year follow-up of a much younger cohort in Ecuador has 19 percent attrition (Araujo, Bosch, and Schady, 2018).

[41] This section draws from Molina Millán and Macours (2017), which contains a more detailed explanation of the approach taken and further rationale for the selection correction.

[42] The few variables that are not balanced in the final 2010 sample (column 5 and 9) are similarly not balanced on the full baseline target sample (column 1).

the weights, we estimate the probability of being found among those tracked during the intensive phase.

We calculate attrition-correction weights separately for each survey instrument. A large number of socio-economic variables observed in or calculated from the program census was considered for predicting attrition, informed in particular by the nature of migration from the regions. These include all of the baseline variables shown in the balance table for the sample in appendix Table A1, capturing individual-, household- (parental education, demographics, economic activities, and assets), and locality-level characteristics. As connectivity could be a good predictor of tracking success, we include two variables to capture the social network of the individual (village size and family network size), some more detailed household structure variables, and a set of proxy variables meant to capture the possible temporary nature of some households' residence in the village at baseline.[43] We similarly consider locality-level characteristics that could be push or pull factors for migration: remoteness (measured using distance to night light and altitude), location in a coffee producing area, and having been affected by hurricane Mitch, a severe storm in the area in 1998. Finally, as further proxy measures for locations with a concentration of more temporary workers, we introduce two variables capturing the level of attrition between the program census and the first baseline survey (i.e., between May and August of 2000) in the locality of origin of the individual: the share of individuals attrited, and whether any individual in the target age cohort attrited.[44] Individuals from such locations not only were more likely to attrit, but also could be harder to trace, as contacts with the community of origin could be limited.

Because of the large number of potential variables to consider and because there are relatively few individuals not found after intensive tracking, we follow Doyle et al. (2017) to select a reduced set of predictors. Separate estimations are carried out to model attrition for the household and individual instruments. First, we first estimate bivariate regressions in which each potential predictor is examined to determine whether a significant difference exist between the means for those found and not found during intensive tracking. All estimates use the survey sample weights and standard errors are clustered at the locality level. This testing is conducted separately for the early and late treatment groups. Results are shown in first four columns of appendix Table F4 for attrition in the household instrument and in final four columns for the individual instrument. The correlates of having been found during intensive tracking differ between treatment groups and also between the survey instruments.

We retain as potential predictors all indicators found to be statistically significantly different for the early or late treatment group. We then carry out a first estimate the probability of having been found during intensive tracking on this set of baseline predictor variables for each treatment group, using the sample of those not found during regular tracking. To account for collinearity between measures, the baseline predictor set is restricted further by conducting stepwise selection of variables with backward elimination and using the adjusted R^2 as the information criterion. Strata and regional fixed effects, as well as 6-month age dummies are included as fixed predictors in all models.

In the final step, we estimate the probability of having been found during the intensive tracking phase for both early and late treatment group together, keeping only the predictors as

[43] In particular, we include a set of indicators as proxies for whether the household comprised temporary workers on one of the large coffee plantations (haciendas). These households were captured in the program census but likely were not permanent residents.

[44] The baseline survey was conducted shortly after the public lottery and before the start of the transfers.

indicated by the stepwise procedure, as well as the strata and regional fixed effects and the 6-month age dummies, all interacted with the treatment variable. In addition, we add a set of field supervisor fixed effects, to capture potential differences between survey teams in effectiveness in tracking. The resulting regression has good predictive power (the linear probability model has an R^2 of 51 percent for household level attrition and 37 percent for individual level attrition). Table F5 shows the linear probability model estimates for each survey.

The probability of having been found during intensive tracking (conditional on not having been found during regular tracking) estimated via a probit for this last specification shown in Table F5 is then used to determine the weights for the attrition selection. All observations found during regular tracking are assigned a weight of 1, while those found during the intensive tracking are assigned a weight 1/Prob (found during intensive tracking | not having been found during regular tracking). Finally, these weights are then multiplied with the sample weights. Final attrition-correction weights vary between 1 and 35 for the household instrument, and 1 and 86 for the individual instrument, with corresponding averages of 3.4 and 3.9, respectively.[45]

We overweight individuals interviewed during the intensive tracking phase, as these individuals were, by definition, more difficult to find, and therefore more likely to be similar to those not found at all. Empirically, observable characteristics are also much better predictors for the subsample that was intensively tracked than for the full sample, suggesting that selection on observables for this subsample is a more plausible assumption than for the full sample. Nevertheless, we also present results with the more standard inverse probability weighting (with weights estimated using the probability of being found in the entire sample) in Table B1.1, for comparison. Estimations of weights for the standard IPW followed a similar process of covariate selection.

Alternative treatments of attrition relying on different assumptions: no correction, Lee bounds and proxy's

In addition to the two estimates with inverse probability weights, we also present a number of alternative estimations to examine the sensitivity of the findings to different assumptions about attrition. First, because attrition rates and baseline characteristics of the interviewed samples are balanced across treatment groups, we re-estimate the main results without any attrition corrections (appendix B, Tables B2.1-B2.4). These estimates rely on the assumption that balance on observables translates into balance on unobservables, and leads to internally valid estimates. With treatment effect heterogeneity, however, they may have more limited external validity as they are less likely to be representative for the population that was not found. Results from that exercise demonstrate that the main results are not driven by the attrition correction and point in the same direction for all of the families of outcomes (as also summarized in panel E of Table B1.1). Give the relatively low levels of attrition, it is indeed intuitive that results are very similar. Second, we estimate Lee bounds (Table B1.2). These estimates rely on the monotonicity assumption, implying that assignment to the treatment group can affect attrition in only one direction. Because attrition in this context may be correlated both with program compliance and with labor market and migration outcomes (as discussed above), and each of those could be affected by early and late treatment exposure in various ways, the validity of the monotonicity

[45] With the exception of a few outliers, the distribution of weights is not highly skewed with 97 percent of household and 96 percent of individual weights less than nine. Only two observations have individual weights higher than 32 and omitting these two observations from the analysis does not alter any of the findings. With the exception of those 2 outliers, the distribution of weights is similar when using conventional IPW estimates.

assumption in this setting is unclear (but presented for completeness). Third, for selected available outcomes we present results using proxy measures, which leads to much lower rates of missing information, but also likely introduces measurement error in the outcome variables (Table A3).

TABLE F1. ATTRITION AND TRACKING IN THE EARLY AND LATE TREATMENT GROUPS

Panel A	**Probability of having been interviewed: Household instrument**			
	(1)	(2)	(3)	(4)
	Found (i.e., interviewed)	Found during regular tracking	Found during intensive tracking (intensive tracking subsample)	Found (incorporating proxy information)
ITT	-0.014	-0.023	-0.020	-0.025
	(0.027)	(0.037)	(0.075)	(0.021)
Mean late treatment	0.905	0.751	0.62	0.969
Observations	1138	1138	297	1138

Panel B	**Probability of having been interviewed: Individual instrument**		
	(1)	(2)	(3)
	Found (i.e., interviewed)	Found during regular tracking	Found during intensive tracking (intensive tracking subsample)
ITT	0.008	0.017	0.004
	(0.037)	(0.044)	(0.071)
Mean late treatment	0.793	0.455	0.62
Observations	1138	1138	611

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Boys age 9-12 at the start of the program in November 2000.

TABLE F2. COMPARISON OF BASELINE BALANCE BY TRACKING STATUS AND 2010 SURVEY INSTRUMENT

	Baseline Sample: Cohort 9-12 Boys		Household Instrument Outcomes				Individual Instrument Outcomes			
			Found During Regular Tracking		Final Sample		Found During Regular Tracking		Final Sample	
	N=1138		N=826		N=1006		N=527		N=907	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Individual Characteristics										
Age at start of transfer in months	-0.046*	(0.026)	-0.086*	(0.047)	-0.068**	(0.032)	-0.129*	(0.068)	-0.052	(0.041)
No grades attained (=1)	-0.008	(0.064)	-0.025	(0.065)	-0.014	(0.063)	-0.018	(0.073)	-0.016	(0.065)
Highest grade attained	0.071	(0.16)	0.136	(0.16)	0.094	(0.15)	0.173	(0.18)	0.111	(0.16)
Worked in last week (=1)	-0.054*	(0.027)	-0.064**	(0.029)	-0.064**	(0.030)	-0.039	(0.038)	-0.060**	(0.029)
Participated in some economic activity (=1)	-0.009	(0.035)	-0.016	(0.037)	-0.014	(0.035)	0.009	(0.042)	-0.008	(0.037)
Household Characteristics: Education										
Distance to nearest school (minutes)	0.744	(4.76)	2.547	(4.56)	0.594	(4.54)	4.308	(5.91)	1.153	(4.31)
Household head no grades attained (=1)	0.005	(0.031)	0.005	(0.040)	0.020	(0.035)	0.024	(0.034)	0.036	(0.036)
Household head 3 plus grades attained (=1)	0.025	(0.029)	0.046	(0.031)	0.032	(0.027)	0.060*	(0.033)	0.033	(0.026)
Mother no grades attained (=1)	-0.046	(0.037)	-0.097**	(0.042)	-0.053	(0.043)	-0.100*	(0.053)	-0.040	(0.045)
Mother 3 plus grades attained (=1)	0.072	(0.044)	0.111***	(0.038)	0.068	(0.041)	0.171***	(0.052)	0.057	(0.042)
Household Characteristics: Demographics										
Father not living in same household (=1)	0.017	(0.031)	0.041	(0.029)	0.013	(0.027)	0.092**	(0.040)	0.011	(0.025)
Mother not living in same household (=1)	0.011	(0.017)	0.027	(0.017)	0.013	(0.017)	0.022	(0.018)	0.015	(0.019)
Child of household head (=1)	-0.016	(0.024)	-0.061**	(0.024)	-0.020	(0.023)	-0.074**	(0.031)	-0.023	(0.024)
Number of children of household head	-0.241	(0.22)	-0.511**	(0.23)	-0.326	(0.23)	-0.487*	(0.29)	-0.311	(0.23)
Female household head (=1)	0.030	(0.020)	0.035*	(0.018)	0.018	(0.018)	0.053*	(0.026)	0.019	(0.018)
Age of household head	0.339	(0.87)	0.556	(0.87)	0.473	(0.79)	0.290	(1.11)	0.843	(0.85)
Number of household members	-0.047	(0.19)	-0.205	(0.24)	-0.159	(0.22)	0.028	(0.27)	-0.089	(0.22)
Nuclear household (=1)	-0.017	(0.040)	-0.023	(0.044)	-0.019	(0.043)	-0.071	(0.052)	-0.038	(0.044)
Multigenerational household (=1)	-0.037	(0.035)	-0.005	(0.034)	-0.020	(0.037)	0.023	(0.039)	-0.009	(0.038)

Other household structure (=1)	0.054*	(0.028)	0.028	(0.031)	0.038	(0.028)	0.047	(0.038)	0.047	(0.032)
Number of children aged 0-8	0.021	(0.11)	-0.067	(0.12)	-0.097	(0.11)	0.095	(0.14)	-0.087	(0.12)
Number of children age 9 to 12	-0.036	(0.058)	-0.078	(0.070)	-0.072	(0.062)	-0.087	(0.076)	-0.066	(0.068)
Household Characteristics: Economic Activities & Assets										
Household head main occupation is agric. (=1)	-0.021	(0.038)	-0.003	(0.046)	-0.018	(0.041)	0.004	(0.049)	-0.018	(0.042)
Size of landholdings (1000 square meters)	-1.987	(1.85)	-2.822	(2.02)	-2.500	(1.80)	-3.730	(2.28)	-1.656	(1.82)
Log of size of landholdings	-0.270	(0.41)	-0.049	(0.43)	-0.087	(0.42)	0.144	(0.60)	0.013	(0.44)
Number of parcels of land	-0.036	(0.073)	-0.018	(0.076)	-0.008	(0.074)	0.010	(0.094)	0.017	(0.073)
Log predicted expenditures (per capita)	-0.009	(0.022)	0.029	(0.026)	0.003	(0.025)	0.009	(0.034)	0.001	(0.025)
Wealth index - housing characteristics	0.200	(0.16)	0.328**	(0.16)	0.223	(0.15)	0.226	(0.17)	0.159	(0.15)
Wealth index - productive assets	-0.258**	(0.11)	-0.209*	(0.12)	-0.263**	(0.11)	-0.188	(0.15)	-0.201*	(0.11)
Wealth index - other assets	-0.040	(0.18)	-0.127	(0.19)	-0.059	(0.18)	-0.123	(0.21)	-0.067	(0.18)
Number of rooms in house	0.030	(0.074)	0.067	(0.086)	0.051	(0.074)	0.022	(0.089)	0.021	(0.075)
Cement block walls (=1)	0.044	(0.043)	0.065	(0.044)	0.036	(0.043)	0.043	(0.058)	0.007	(0.042)
Zinc roof (=1)	0.050	(0.085)	0.020	(0.082)	0.052	(0.083)	0.020	(0.090)	0.057	(0.085)
Tile roof (=1)	-0.093	(0.097)	-0.047	(0.097)	-0.094	(0.095)	-0.042	(0.10)	-0.098	(0.096)
Dirt floor (=1)	-0.030	(0.036)	-0.049	(0.038)	-0.039	(0.036)	-0.017	(0.045)	-0.016	(0.038)
Latrine or toilet (=1)	0.035	(0.063)	0.083	(0.067)	0.062	(0.065)	0.103	(0.071)	0.073	(0.065)
Electric light (=1)	0.074	(0.050)	0.116**	(0.057)	0.084	(0.052)	0.120**	(0.059)	0.078	(0.054)
Radio (=1)	0.020	(0.034)	0.010	(0.040)	0.005	(0.040)	-0.027	(0.042)	0.004	(0.041)
Work animals (=1)	-0.0541*	(0.029)	-0.026	(0.030)	-0.0573*	(0.031)	-0.046	(0.043)	-0.048	(0.031)
Fumigation sprayer (=1)	-0.073	(0.047)	-0.062	(0.058)	-0.073	(0.050)	-0.027	(0.056)	-0.053	(0.050)
Village Characteristics										
Village affected by hurricane Mitch (=1)	-0.062	(0.050)	-0.062	(0.051)	-0.059	(0.053)	-0.016	(0.057)	-0.039	(0.049)
Altitude of village	-21.10	(34.6)	-25.69	(35.9)	-22.51	(35.9)	-26.79	(37.0)	-24.49	(36.9)
Village in coffee producing area (=1)	-0.018	(0.065)	-0.021	(0.066)	-0.010	(0.066)	-0.008	(0.065)	-0.008	(0.066)
Distance to night light (meters)	2276	(2642)	1291	(2619)	1880.0	(2607)	2730	(2758)	1980	(2645)
Live in Tuma region (=1)	0.193	(0.15)	0.155	(0.14)	0.194	(0.14)	0.162	(0.15)	0.187	(0.14)
Live in Madriz region (=1)	0.017	(0.12)	0.044	(0.12)	0.027	(0.13)	0.021	(0.12)	0.017	(0.13)
Social Capital										
Family network size (individuals)	23.52**	(10.3)	27.65**	(12.1)	25.94**	(10.9)	19.20	(14.1)	26.17**	(11.7)
Population size village	257.6***	(88.6)	249.0**	(94.6)	257.2***	(88.7)	250.1**	(101)	242.1***	(86.9)

Proxy's of Permanent Residence in Village										
Own house (=1)	-0.052	(0.043)	-0.021	(0.046)	-0.064	(0.043)	-0.008	(0.041)	-0.048	(0.041)
House is obtained in exchange for service/labor (=1)	0.019	(0.039)	-0.016	(0.032)	0.023	(0.036)	-0.014	(0.033)	0.011	(0.037)
Address in hacienda (=1)	-0.022	(0.058)	-0.023	(0.058)	-0.004	(0.056)	-0.028	(0.059)	-0.007	(0.054)
Address in hacienda & house rented (=1)	0.032	(0.033)	0.018	(0.031)	0.047	(0.031)	0.017	(0.022)	0.038	(0.028)

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. All estimates control for strata fixed effects. Regressions are weighted to account for sampling providing population estimates. Distance to night light (meters) is linear distance from household to an area with stable night light detected by a satellite (DMSP-OLS Nighttime Lights). House received in exchange for services is an indicator variable for households who received the house in exchange for labor services. Address in hacienda is an indicator for households whose address refers to a location on a large plantation (hacienda).

TABLE F3. RELATIONSHIP BETWEEN THE PROBABILITY OF BEING FOUND, BASELINE COVARIATES, AND TREATMENT

	X		*X*T*		*T*	
Individual Characteristics						
Age at start of transfer in months	-0.024**	(0.011)	-0.017	(0.018)	0.170	(0.20)
No grades attained (=1)	-0.012	(0.029)	-0.056	(0.060)	0.010	(0.037)
Highest grade attained	-0.010	(0.012)	0.019	(0.026)	-0.038	(0.044)
Worked in last week (=1)	0.005	(0.035)	-0.051	(0.057)	-0.007	(0.030)
Participated in some economic activity (=1)	-0.007	(0.024)	-0.018	(0.047)	-0.011	(0.032)
Household Characteristics: Education						
Distance to nearest school (minutes)	0.000	(0.000)	0.000	(0.000)	-0.018	(0.033)
Household head no grades attained (=1)	-0.039	(0.031)	0.041	(0.054)	-0.037	(0.040)
Household head 3 plus grades attained (=1)	-0.006	(0.035)	0.036	(0.051)	-0.026	(0.033)
Mother no grades attained (=1)	0.100**	(0.040)	-0.031	(0.064)	0.002	(0.052)
Mother 3 plus grades attained (=1)	-0.008	(0.027)	-0.016	(0.069)	-0.009	(0.035)
Household Characteristics: Demographics						
Father not living in same household (=1)	-0.099*	(0.050)	-0.017	(0.064)	-0.011	(0.027)
Mother not living in same household (=1)	-0.12	(0.088)	0.059	(0.13)	-0.020	(0.031)
Child of household head (=1)	0.057	(0.057)	-0.012	(0.088)	-0.004	(0.083)
Number of children of household head	0.019***	(0.005)	-0.014	(0.011)	0.055	(0.068)
Female household head (=1)	-0.046	(0.051)	-0.065	(0.071)	-0.007	(0.027)
Age of household head	0.002	(0.002)	0.000	(0.0026)	-0.027	(0.13)
Number of household members	0.011	(0.007)	-0.016	(0.011)	0.120	(0.099)
Nuclear household (=1)	0.038	(0.031)	-0.010	(0.051)	-0.009	(0.043)
Multigenerational household (=1)	-0.029	(0.033)	0.106*	(0.053)	-0.042	(0.038)
Other household structure (=1)	-0.035	(0.063)	-0.123	(0.095)	0.006	(0.026)
Number of children aged 0-8	0.018	(0.012)	-0.058**	(0.022)	0.108*	(0.055)
Number of children age 9 to 12	0.023	(0.015)	-0.071*	(0.039)	0.111	(0.068)
Household Characteristics: Economic Activities&Assets						
Household head main occupation is ag. (=1)	0.063	(0.047)	-0.008	(0.074)	-0.008	(0.075)
Size of landholdings ('000 sq meters)	0.000	(0.000)	-0.001	(0.002)	0.006	(0.044)
Log of size of landholdings	0.004	(0.005)	0.011	(0.010)	-0.100	(0.091)
Number of parcels of land	0.030	(0.020)	0.063	(0.043)	-0.072	(0.059)
Log predicted expenditures (pc)	-0.121*	(0.066)	0.134	(0.089)	-1.055	(0.68)

Wealth index - housing characteristics	-0.024*	(0.014)	0.017	(0.020)	-0.014	(0.029)
Wealth index - productive assets	0.034**	(0.015)	0.001	(0.023)	-0.007	(0.028)
Wealth index - other assets	-0.011	(0.014)	-0.027	(0.019)	-0.016	(0.026)
Village Characteristics						
Village affected by hurricane Mitch (=1)	-0.096***	(0.024)	0.0935*	(0.048)	-0.101**	(0.038)
Altitude of village ('000 meters)	0.027	(0.14)	-0.098	(0.20)	0.046	(0.13)
Village in coffee producing area (=1)	-0.004	(0.036)	0.003	(0.060)	-0.018	(0.058)
Distance to night light ('000 meters)	-0.003	(0.002)	0.000	(0.002)	0.062*	(0.036)
Live in Tuma region (=1)	-0.123***	(0.032)	-0.013	(0.042)	0.017	(0.018)
Live in Madriz region (=1)	0.0375	(0.028)	0.0908**	(0.037)	-0.033	(0.032)
Social Capital						
Family network size (individuals) '000	0.678***	(0.22)	0.094	(0.27)	-0.040	(0.042)
Population size village '000	-0.013	(0.027)	0.003	(0.042)	-0.014	(0.042)
Proxy's of Permanent Residence in Village						
Own house (=1)	0.146**	(0.064)	-0.080	(0.091)	0.057	(0.082)
House is obtained in exchange for service/labor (=1)	-0.154**	(0.071)	0.034	(0.13)	-0.015	(0.029)
Address in hacienda (=1)	-0.111	(0.072)	0.076	(0.099)	-0.028	(0.030)
Address in hacienda & house rented (=1)	-0.289**	(0.14)	0.231	(0.16)	-0.026	(0.029)

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and in parentheses. Boys 9-12 years old in November 2000. For each baseline variable X, the table presents coefficient estimates of the linear probability model: $Prob(found) = F(X, T, XT)$.

TABLE F4. CORRELATES OF THE PROBABILITY OF BEING FOUND DURING THE INTENSIVE TRACKING PHASE

	Household Instrument				Individual Instrument			
	Treatment N=160		Control N=137		Treatment N=311		Control N=300	
	Coef. (1)	SE (2)	Coef. (3)	SE (4)	Coef. (5)	SE (6)	Coef. (7)	SE (8)
Individual Characteristics								
Age at start of transfer in months	-0.446**	(0.18)	-0.382	(0.23)	-0.375***	(0.12)	-0.421***	(0.091)
No grades attained (=1)	-0.122	(0.16)	-0.071	(0.12)	-0.209**	(0.084)	-0.147**	(0.055)
Highest grade attained	0.042	(0.44)	-0.074	(0.37)	0.220	(0.29)	0.054	(0.18)
Worked in last week (=1)	-0.038	(0.066)	0.043	(0.084)	-0.029	(0.052)	0.047	(0.047)
Participated in some economic activity (=1)	-0.025	(0.076)	0.003	(0.069)	-0.009	(0.058)	0.029	(0.048)
Household Characteristics: Education								
Distance to nearest school (minutes)	-4.960**	(2.33)	1.910	(7.59)	-11.18**	(4.85)	-8.546	(5.25)
Household head no grades attained (=1)	0.067	(0.13)	-0.114	(0.11)	0.030	(0.11)	-0.128**	(0.061)
Household head 3 plus grades attained (=1)	0.001	(0.068)	-0.011	(0.088)	0.049	(0.054)	0.054	(0.068)
Mother no grades attained (=1)	0.323**	(0.14)	0.220	(0.14)	0.143	(0.086)	0.027	(0.073)
Mother 3 plus grades attained (=1)	-0.177	(0.15)	0.074	(0.099)	-0.115	(0.091)	0.112*	(0.056)
Household Characteristics: Demographics								
Father not living in same household (=1)	-0.190**	(0.068)	-0.008	(0.12)	-0.179***	(0.052)	-0.034	(0.057)
Mother not living in same household (=1)	-0.049	(0.084)	-0.001	(0.085)	0.003	(0.052)	-0.022	(0.029)
Child of household head (=1)	0.123	(0.082)	-0.068	(0.092)	0.058	(0.053)	0.014	(0.032)
Number of children of household head	0.658	(0.56)	0.711	(0.55)	0.134	(0.21)	0.302	(0.28)
Female household head (=1)	-0.133**	(0.051)	0.025	(0.063)	-0.083**	(0.035)	0.017	(0.040)
Age of household head	4.009	(2.69)	4.487*	(2.56)	2.987*	(1.51)	-0.579	(1.86)
Number of household members	-0.134	(0.73)	1.054	(0.67)	-0.472	(0.33)	0.078	(0.39)
Nuclear household (=1)	0.002	(0.11)	-0.022	(0.11)	0.062	(0.070)	0.102	(0.090)
Multigenerational household (=1)	0.169*	(0.085)	0.065	(0.091)	0.016	(0.084)	-0.069	(0.078)
Other household structure (=1)	-0.171*	(0.088)	-0.043	(0.052)	-0.077	(0.063)	-0.033	(0.048)
Number of children aged 0-8	-0.748	(0.59)	0.649*	(0.33)	-0.804***	(0.23)	0.107	(0.16)
Number of children age 9 to 12	-0.183	(0.20)	0.167	(0.12)	-0.110	(0.13)	0.023	(0.087)
Household Characteristics: Economic Activities & Assets								

Household head main occupation is ag. (=1)	0.018	(0.13)	0.080	(0.063)	0.017	(0.065)	0.0493	(0.059)
Size of landholdings (1000 square meters)	-5.661	(10.7)	-2.630	(2.46)	-2.683	(5.69)	-5.810**	(2.29)
Log of size of landholdings	1.342	(1.68)	-0.055	(0.55)	1.394	(0.96)	0.392	(0.60)
Number of parcels of land	0.288	(0.19)	0.025	(0.080)	0.199	(0.13)	-0.047	(0.084)
Log predicted expenditures (per capita)	-0.070	(0.079)	-0.116	(0.10)	0.007	(0.051)	-0.040	(0.056)
Wealth index - housing characteristics	-0.402	(0.42)	-0.327	(0.39)	-0.631	(0.41)	-0.338	(0.22)
Wealth index - productive assets	0.180	(0.21)	0.439**	(0.18)	0.372*	(0.19)	0.117	(0.12)
Wealth index - other assets	-0.164	(0.17)	-0.277	(0.20)	-0.484***	(0.17)	-0.287*	(0.14)
Village Characteristics								
Village affected by hurricane Mitch (=1)	-0.022	(0.065)	-0.106*	(0.054)	0.0339	(0.073)	-0.051*	(0.026)
Altitude of village	-16.20	(44.1)	-16.76	(52.2)	-0.0909	(0.081)	-0.049	(0.054)
Village in coffee producing area (=1)	0.025	(0.094)	-0.071	(0.081)	-27.87	(38.4)	-5.22	(22.7)
Distance to night light (meters)	-3017**	(1288)	-1407	(1707)	-5302***	(1486)	-2108	(1512)
Live in Tuma region (=1)	-0.161**	(0.068)	-0.339***	(0.098)	-0.387***	(0.076)	-0.335***	(0.098)
Live in Madriz region (=1)	0.106	(0.061)	0.087	(0.063)	0.160**	(0.074)	0.086	(0.056)
Social Capital								
Family network size (individuals)	31.65**	(11.5)	17.98**	(6.86)	41.72***	(11.3)	15.41**	(7.01)
Population size village	-3.007	(80.2)	-40.26	(52.0)	-142.6**	(68.0)	-68.24	(43.1)
Proxy's of Permanent Residence in Village								
Own house (=1)	-0.046	(0.13)	0.230***	(0.078)	0.089	(0.083)	0.110	(0.074)
House is obtained in exchange for service/labor (=1)	0.062	(0.12)	-0.114**	(0.052)	-0.094	(0.085)	-0.073	(0.055)
Address in hacienda (=1)	0.081	(0.11)	-0.128	(0.10)	-0.020	(0.067)	-0.096	(0.085)
Address in hacienda & house rented (=1)	0.065	(0.12)	-0.172	(0.10)	-0.042	(0.069)	-0.094	(0.060)
Variables Indicating Very Early Attrition								
Probability of attrition prior to program start in locality	-0.005	(0.012)	-0.034	(0.028)	-0.032**	(0.013)	-0.047*	(0.025)
Nobody of target sample attrited before program start	0.077	(0.078)	0.221**	(0.090)	0.283***	(0.085)	0.188**	(0.082)

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses.

TABLE F5. LINEAR PROBABILITY ESTIMATES FOR PROBABILITY OF BEING FOUND DURING INTENSIVE TRACKING PHASE

	Household Instrument		Individual Instrument	
	Coef. (SE)	Coef. ET Interaction (SE)	Coef (SE)	Coef. ET Interaction (SE)
Early treatment (ET)=1		-0.581		-0.347
		(0.38)		(0.30)
No grades attained (=1)			-0.097	-0.014
			(0.073)	(0.12)
Distance to nearest school (minutes)			0.000	-0.001
			(0.001)	(0.001)
Mother no grades attained (=1)	0.159*	0.022		
	(0.094)	(0.13)		
Mother 3 plus grades attained (=1)			-0.032	-0.040
			(0.061)	(0.078)
Household head no grades attained (=1)			-0.103*	0.140*
			(0.054)	(0.082)
Father not living in same household (=1)	-0.144	0.134		
	(0.17)	(0.19)		
Child of household head (=1)	0.0501	0.209		
	(0.18)	(0.22)		
Female household head (=1)	0.374***	-0.506***	0.039	-0.196*
	(0.12)	(0.18)	(0.088)	(0.10)
Other household structure (=1)	0.426***	-0.434***		
	(0.12)	(0.15)		
Number of children aged 0-8	0.072***	-0.131***	0.008	-0.058**
	(0.022)	(0.036)	(0.017)	(0.022)
Number of parcels of land	-0.168**	0.303***		
	(0.066)	(0.092)		
Size of landholdings ('000 sq meters)			-0.002	0.003
			(0.002)	(0.002)
Wealth index - housing characteristics			-0.035	0.007
			(0.023)	(0.033)
Wealth index - productive assets	0.058	-0.033	-0.005	0.046
	(0.035)	(0.060)	(0.033)	(0.056)
Village affected by hurricane Mitch (=1)	-0.278**	0.163	-0.104*	0.208*
	(0.12)	(0.15)	(0.053)	(0.11)
Distance to night light (km)			0.007*	-0.009
			(0.004)	(0.007)
Live in Tuma region (=1)	-0.283***	0.379***	-0.276***	0.246
	(0.082)	(0.11)	(0.079)	(0.15)

Live in Madriz region (=1)	0.179	-0.080	0.162**	-0.257**
	(0.13)	(0.19)	(0.070)	(0.10)
Population size village	0.001	0.001	0.000	0.001*
	(0.001)	(0.001)	(0.000)	(0.001)
Own house (=1)	0.371**	-0.331*		
	(0.15)	(0.18)		
House is obtained in exchange for service/labor (=1)	0.407***	-0.356*		
	(0.12)	(0.21)		
Address in hacienda & house rented (=1)	-0.231	0.427	-0.204	0.306*
	(0.23)	(0.27)	(0.16)	(0.18)
Probability of attrition prior to program start in comarca			-0.133	-0.475
			(0.31)	(0.97)
Nobody of target sample attrited before program start	0.120	-0.075	0.043	0.006
	(0.11)	(0.15)	(0.11)	(0.16)
Age fixed effects	YES	YES	YES	YES
Strata fixed effects	YES	YES	YES	YES
Supervisor fixed effects	YES		YES	
Observations	297		611	
R-squared	0.51		0.37	

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and in parentheses. First column shows coefficient of variable alone, second column coefficient of the variable interacted with the early treatment dummy.

APPENDIX G: SPILLOVERS AND FAMILY NETWORKS

Spillovers—One reason the experimental differential estimates may underestimate absolute treatment effects is that there could be positive spillovers between the early and late treatment groups. As detailed in the description of the evaluation (Section II.B), a set of 42 rural localities in six municipalities were randomized into early or late treatment. Many of these shared geographic borders (with no buffer zone), including some with the opposite treatment group, opening up the possibility of spillover effects. In this appendix, we describe the methodology we use to test for spillover effects exploiting GPS information on initial household location.

In contrast to a more typical treatment and control design, in which one might expect spillovers in one direction only (positive or negative), the phase-in design of early and late treatment implies that the direction of spillovers coult vary over time. For example, if there were positive spillovers from treatment to non-treatment areas during program operation, then there could have been spillovers from early to late treatment areas in the first years of the program when only early treatment households received transfers, followed by spillovers from late to early treatment areas in later years when only late treatment households received transfers. Even if all spillover effects were positive, the net effect of spillovers on early treatment (measured after 10 years), and therefore on the differential between early and late treatment, would be ambiguous.

To explore the potential influence of spillover effects on the main differential findings, we exploit GPS location data for sample households and analyze whether differential program effects differ depending on whether households were living in close proximity to many other households receiving the program during early treatment.[46] To do so, we use the variation in the density of households with early treatment in a 3-kilometer radius around each household, which results from the locality-level randomization, and the fact that the geographical borders of the locality (based on census segments) do not necessarily correspond to village borders. The density itself is likely to capture other unobserved characteristics of the environment, but the interaction of the density with early treatment provides an estimate that accounts for both spillovers, as well as for any potential multiplier effects among early treatment households. The possibility of the latter, in areas where treatment density was high, is consistent with evidence from contemporaneous qualitative work that demonstrated strong united enthusiasm for improved schooling during the CCT, evidenced by community mobilization to ensure additional teachers if needed (Adato and Roopnaraine 2004).

Results in appendix Table G1 suggest that treatment effects appear to be larger when early treatment density is higher but only one of the interactions, for the socio-emotional family, is significant at 10 percent. If anything, the signs of the interaction effects do not suggest differential effect estimates are underestimated due to spillovers, and they are instead consistent with the possibility of a positive multiplier effect within the early treatment group.

Family Kinship Networks—Family networks have been shown to be important for anti-poverty programs in many contexts. Moreover, heterogeneous treatment effects related to such networks can shed light on the mechanisms underlying impacts. Building on the literature examining the importance of family and broader kinship networks for impacts of CCTs in the

[46] Initial baseline GPS locations are available only for households targeted for 2010 follow-up, and not for all households in the 2000 program census. We use the observed program density in the sample as a proxy for actual program density in the population.

short run (Angelucci and Di Giorgio 2009; Angelucci et al. 2010), we analyze to what extent program effects differ depending on the strength of pre-existing family networks.

The program census carried out in May 2000 included all households in both early and late treatment localities, and administered individual rosters for each household. In addition to the demographic and schooling information for individuals, for the purpose of program administration the census also recorded and digitized first and last names of all household members. In many parts of Latin America including Nicaragua, individuals have two surnames and naming conventions follow the pattern in which a child typically takes the first surname of his father as his first surname, and the first surname of his mother as his second surname. For example, if the paternal surname is Godoy Sandino (F1, f1) and the maternal surname is Darío Martí (F2, f2), then their child's surname would be Godoy Darío (F1, F2).

In the spirit of the approach taken for constructing family networks in Mexico in analyses of *PROGRESA* (Angelucci and De Giorgi 2009; Angelucci et al. 2010; Angelucci, De Giorgi, and Rasul 2017), we use this naming convention to construct different measures of the potential family network in each village based on surname matches. Specifically, we use the program census data to determine, for each boy, the total number of other individuals living in the same village[47] who have at least one surname in common with that boy, regardless of its position (i.e., first or second surname). Alternatives to this definition, for example requiring the same position (first or second) or using the name of the household head instead, lead to quantitatively smaller networks that are highly correlated with our broader definition. While we cannot verify that all individuals within the family network as defined here would have a direct familial relationship, many likely do (including siblings and half-siblings, cousins, grandfathers, and aunts/uncles). Median family network size in these rural communities is nearly 60 persons (average 83) and the median family network as a share of the total number of individuals in the community is nearly 20 percent (average 26 percent). We transform the family network size into a binary variable indicating above and below median network size, in part to mitigate concerns about measurement error for those with the same name who do not have direct family relationships (Angelucci et al. 2010).

We use these additional variables to analyze heterogeneous treatment effects. Results for grades attained and the families of outcomes presented in the paper are shown in appendix Table G2. With the additional variables, the interpretation on the treatment variable is now the ITT differential impact for boys with below median family network size at baseline and the interaction captures differences in that impact for those with initially high family network size. There is little evidence of heterogeneous treatment effects, though for the education and socio-emotional families the point estimates hint at somewhat smaller effects on those outcomes for boys with initially larger family networks. Assuming the lack of differential effects is not due only to low power,[48] we conclude that the long-term program effects were not concentrated among those with initially stronger or weaker family networks.

[47] We use village, as defined in the 1995 National Census, as the relevant geographic area rather than the larger locality area used for randomization and clustering in the evaluation, as the former constitutes a more concentrated settlement and is arguably a more natural geographic area in which to consider network effects.

[48] For comparison, Angelucci et al. (2010) have a male sample four times as large.

TABLE G1: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, BY EARLY TREATMENT DENSITY

	Education		Learning Family Z-Score	Economic Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Absolute (5% Trim)	Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ITT	0.237	0.160	0.039	0.022	-0.247	-0.305	-0.071
	(0.494)	(0.118)	(0.197)	(0.241)	(0.161)	(0.187)	(0.135)
Early treatment density	-0.789	-0.258**	-0.223	0.093	0.085	0.233	-0.318**
	(0.526)	(0.123)	(0.203)	(0.203)	(0.227)	(0.214)	(0.155)
ITT * Early treatment density	0.673	0.124	0.350	0.238	0.481	0.434	0.399*
	(0.801)	(0.194)	(0.333)	(0.327)	(0.289)	(0.290)	(0.215)
Observations	1,006	1,007	907	1,006	1,006	997	900
R-squared	0.426	0.340	0.450	0.148	0.104	0.107	0.156

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and are given in parentheses. Early treatment density is defined as the fraction of early treatment households within a 3 km radius. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Z-scores are calculated as the mean and divided by the standard deviation of the late-treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Variables in columns 1-2 and 4-6 measured using the 2010 household survey; variables in columns 3 and 7 measured using the 2010 individual survey.

TABLE G2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, BY FAMILY NETWORK SIZE

	Education		Learning Family Z-Score	Economic Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Absolute (5% Trim)	Absolute (5% Trim)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ITT	0.295	0.127*	0.147	0.260**	0.185*	0.177**	0.096
	(0.226)	(0.069)	(0.089)	(0.099)	(0.094)	(0.085)	(0.063)
Family network (>median =1)	-0.023	-0.029	0.110	0.137	0.001	-0.033	0.120*
	(0.270)	(0.065)	(0.081)	(0.105)	(0.143)	(0.127)	(0.066)
ITT * Family network (>median =1)	-0.005	-0.040	0.025	-0.026	0.015	0.038	-0.116
	(0.344)	(0.089)	(0.105)	(0.133)	(0.168)	(0.164)	(0.095)
Observations	1,006	1,007	907	1,006	1,006	997	900
R-squared	0.425	0.339	0.452	0.149	0.097	0.097	0.155

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and are given in parentheses. The late treatment mean for grades attained is 5.5. Family network size is the number of individuals with common surnames in the village. Regressions are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Z-scores are calculated as the mean and divided by the standard deviation of the late-treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Variables in columns 1-2 and 4-6 measured using the 2010 household survey; variables in columns 3 and 7 measured using the 2010 individual survey.

APPENDIX H: NON-EXPERIMENTAL ABSOLUTE PROGRAM EFFECTS —MATCHING

To estimate the absolute effects of the program, we compare 2010 outcomes for boys ages 9–12 at baseline in the early treatment group to the 2010 outcomes for the same cohort of boys in a non-experimental comparison group. In 2002, prior to the phase-in of the late treatment group, 21 non-experimental comparison localities from neighboring rural municipalities were added to enhance the potential for an evaluation of the longer-term effects of the program. The principal criteria for selection included: 1) the same marginality index score cut-offs from the Nicaraguan national census used in the selection of the original 42 localities; 2) minimal ongoing or planned development interventions related to the CCT's objectives; and 3) coverage of the geographic regions of the original municipalities.[49] The comparison group was surveyed in 2002, 2004, and 2010 using the same survey instruments as the experimental groups.

As the ex-ante match of the comparison area to the program areas on locality-level characteristics may not be sufficient to balance household and individual characteristics, we estimate the absolute effects using five nearest neighbors matching (NN5), two nearest neighbors matching (NN2), and non-parametric kernel and local linear matching. We draw on the relatively rich data available and include household and individual characteristics in the propensity score. The matching aims to balance these observable characteristics. To satisfy the unconfoundedness assumption we must assume balance of the unobservables (Rosenbaum and Rubin 1983), as required with any non-experimental method.

The nearest neighbor matching estimators are bias adjusted (Abadie and Imbens 2006; Imbens 2015) and standard errors are clustered at the locality level using the analytic asymptotic variance estimator developed by Abadie and Imbens (2008, 2011) that accounts for the fact that the propensity score is estimated. For kernel and local linear estimates, the standard errors are bootstrapped and the bandwidth is set to be small (0.06) to limit the bias (Todd 2007). We estimate average treatment effects on the treated (ATT) which matches all boys who are 9–12 at the start of the program in the early treatment group with same aged boys in the comparison group. For more direct comparison with the ITT experimental differential results, the early treatment group includes all children who were eligible for treatment, regardless of whether they had taken up the treatment (i.e., it includes the non-compliers).

Estimation of Propensity Score—To estimate the propensity score, we combine data from the 2000 program census, with data from the 2002 household survey for the non-experimental comparison group. There are two important caveats. First, the baseline data used to determine the propensity score for the early treatment and comparison groups are from different years: 2000 for the early treatment group and 2002 for the comparison group. We do not use 2002 data for the early treatment group because this group had already received two years of CCT benefits by 2002. We argue the difference in the timing of the surveys is not likely to be a major source of

[49] More specifically, the comparison sample was drawn from rural municipalities adjacent to or neighboring the six original municipalities. Six comparison municipalities without any major planned development initiatives but with similar levels of poverty and density of schools and health clinics were selected to capture the geographic diversity of the original municipalities. After excluding a small number of localities, the same marginality index used to select the original 42 localities (Arcia 1999; Maluccio 2009), and based on the 1995 Nicaraguan National Census, was calculated for each remaining rural locality. From this exercise, 22 localities with marginality scores in the range targeted by the CCT were identified; one locality that was further way, and thus less likely to be similar, was dropped, leaving 21 comparison localities. A random sample of households was drawn in each. For additional details, see IFPRI (2005).

bias as the value of the variables used in the propensity score are unlikely to have changed much between 2000 and 2002 (e.g., mother's age, mother's and household head's years of education, head's gender, distance to the municipality center).[50]

Second, the data come from different types of survey instruments; census and household surveys.[51] We use the 2000 program census data, rather than the 2000 baseline household survey, in order to include the oversample group in the estimate of the propensity score. The inclusion of the oversample is important for comparability with the differential experimental estimates and also increases the precision of the propensity score estimate. The 2000 program census has a more limited set of variables though all questions in the census and survey instrument are similar for the variables included in the propensity score.

The logit model used to estimate the propensity score is presented in appendix Table H1. We estimate the propensity score using data on boys who are 9–12 at baseline for both early and late treatment groups and the comparison group. While the estimation of the absolute effects does not include the late treatment group, we include it in the construction of the propensity score to increase the sample size and hence the precision of the propensity score estimates.[52] We use all available variables that are similar between the 2000 program census and 2002 household survey and important predictors of either treatment status or the outcomes of interest. We exclude variables whose values are likely to have changed between 2000 and 2002, binary variables which did not have sufficient variation, and information about fathers that was incomplete (e.g., father's age at baseline was missing for more than 20 percent of the sample because it was only asked if the father was a resident of the same household).[53,54] Because of the two-year gap in measurement, it is not possible to consistently measure baseline education variables at the same point in time. As a result, we do not include grades attained or enrollment in the main propensity score, but do include year of birth fixed effects, which are correlated with the education variables.

Propensity Score Balance—We follow Dehejia and Wahba (1999) to determine if the propensity score is balanced across the non-experimental groups and use initial estimates as guides to include interactions or polynomials of variables in the propensity score. We divide the common support into four blocks and test that the propensity score, and each of the variables in the propensity score, are balanced within each block using a t-test.

Appendix Figure H1 presents the distributions of the estimated individual-level propensity score model. Observations above the x-axis are from the early treatment group, and those below from the comparison group. In contrast to what we might see had the groups been randomly allocated, the overlap, while substantial, is imperfect with the treatment group skewed to the

---

[50] When appropriate, comparison group data from 2002 was adjusted in order to be consistent with the data for the early treatment group coming from 2000. For example, age of mother and the child were calculated for the same year, 2000.

[51] An important exception is the locality level marginality index, which is based on the 1995 national census for both the early treatment and the comparison group.

[52] Due to the randomization, the distributions of variables at baseline are similar between early and late treatment groups and using both groups in the estimation of the propensity score does not introduce bias.

[53] We use mother's education from the 2010 survey because it is more complete. In the 2010 survey mother's education was collected for all household members, while in the earlier surveys it was collected for all individuals on the household roster, so mother's education was only available if the mother was living in the same household.

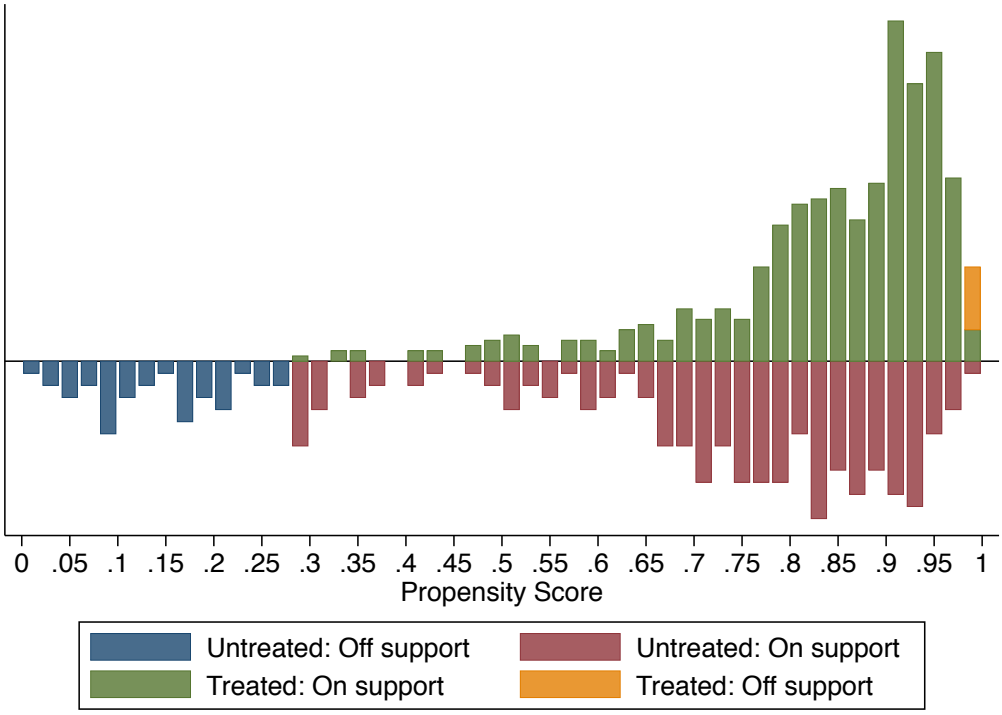
[54] We further do not include the household asset index because not all of the variables are available, however we include the variables that are the same between the two surveys and likely to be correlated with the outcomes.

right and the comparison group to the left. Matching estimators address and correct for this difference in the distributions. To improve the overlap in the covariate distribution, we restrict the analysis to the common support between the experimental group and comparison group. In the main specification reported in the text, we define the common support by trimming all observations that have a propensity score lower than the minimum of the early treatment group distribution, as well as, observations whose propensity score value is greater than the maximum value of the comparison group (i.e. a typical "min-max" common support).

In appendix Table H2 we test the balance of the matching using the "min-max" trim. Columns 1–4 present the p-value on the t-test of the difference in the propensity score and each of the variables used to make the propensity score between the early treatment and comparison group by block. They indicate that the p-score is balanced between the early treatment and comparison group in all blocks, and that only one variable is statistically different within any block at the 5 percent level or lower (less than 1 percent of the block-variable combinations). We further test whether the early treatment and comparison groups are balanced using the NN5 matching estimator on the variables in the propensity score. Appendix Table H2 columns 5 and 6 show that the baseline variables used in the propensity score are balanced between the early treatment and comparison groups; the estimates are close to zero and none are statistically significant at 10 percent level or below.

Absolute Program Effects—The absolute effects using the matching estimators are reported in appendix Table H3. We reproduce the NN5 results from Table 10 in panel A, and then a number of robustness checks to examine the stability of these estimates. First, given the sparse distribution of treatment observations (see appendix Figure H1), we estimate results with the common support defined by trimming all observations that have a propensity score lower than the first and second percentiles (propensity score equal to 0.423 and 0.499, respectively) of the treatment group distribution in panels B and C, rather than the minimum value. Second, we use NN2, non-parametric kernel and local linear matching (panels D, E, and F) to test robustness to different matching estimators. Results are similar in magnitude and significance regardless of which trim or matching estimator is used.

FIGURE H1—INDIVIDUAL-LEVEL PROPENSITY SCORE DISTRIBUTION, BOYS 9–12 IN 2000



Notes: The treated include the early treatment groups and the untreated the comparison group.

TABLE H1— LOGIT RESULTS FOR PROPENSITY SCORE MATCHING AT INDIVIDUAL LEVEL, BOYS 9–12 IN 2000

	Logit (1)	OLS (2)
Age 10 in 2000 (=1)	-0.243	-0.028
	(0.237)	(0.029)
Age 11 in 2000 (=1)	0.549**	0.063**
	(0.243)	(0.027)
Age 12 in 2000 (=1)	-0.171	-0.016
	(0.247)	(0.031)
1995 marginality index	521.454***	81.659***
	(66.590)	(8.046)
1995 marginality index squared	-57.673***	-9.024***
	(7.486)	(0.906)
Mom no education (=1)	0.459**	0.054**
	(0.225)	(0.027)
Mom less than 3 years of education (=1)	15.533*	2.538**
	(8.321)	(1.057)
Mothers age in 2000	-0.017	-0.002
	(0.012)	(0.001)
Household head male (=1)	0.108	0.010
	(0.250)	(0.032)
Household head years of education	-0.186***	-0.025***
	(0.056)	(0.007)
Household head has no education (=1)	-1.081***	-0.130***
	(0.280)	(0.032)
Family size	-0.208	0.003
	(0.180)	(0.014)
Family size squared	0.020*	0.001
	(0.011)	(0.001)
Share of household members age 0-13	1.891***	0.240***
	(0.616)	(0.075)
Has electric light (=1)	-0.744***	-0.096***
	(0.202)	(0.025)
Has work animals (=1)	-0.194	-0.022
	(0.227)	(0.027)
Log (km to the municipality capital)	-0.461***	-0.042**
	(0.168)	(0.019)
Mother no ed. * 1995 marginality index	-3.504*	-0.571**
	(1.873)	(0.237)
R squared		0.203
Observations	1,230	1,230

Notes: *** p<0.01, ** p<0.05, * p<0.10. The dependent variable is the treatment variable for matching, which is 1 if in the early treatment and 0 if in the comparison group.

TABLE H2: BASELINE BALANCE — BY BLOCK AND NN5 MATCHING

	P-value of Difference in Means				NN5	
	Block 1	Block 2	Block 3	Block 4	Diff. in mean	P-value
	(1)	(2)	(3)	(4)	(5)	(6)
Propensity score	0.465	0.536	0.274	0.282		
Age 9 (=1)	0.323	0.924	0.429	0.58	0.042	0.292
Age 10 (=1)	0.318	0.737	0.501	0.271	-0.058	0.134
Age 11 (=1)	0.499	0.668	0.814	0.505	0.039	0.360
Age 12 (=1)	0.565	0.508	0.62	0.419	-0.023	0.613
Marginality Index	0.990	0.542	0.365	0.059	-1.567	0.453
Mother no grades attained (=1)	0.252	0.352	0.281	0.608	-0.038	0.456
Mother 3 plus grades attained (=1)	0.460	0.108	0.652	0.896	-0.050	0.404
Mother's age	0.760	0.111	0.420	0.009	0.157	0.878
Household head male (=1)	0.293	0.244	0.465	0.642	0.003	0.914
Household head years of education	0.676	0.968	0.039	0.131	0.112	0.665
Household head no grades attained (=1)	0.855	0.912	0.052	0.434	-0.037	0.601
Family size	0.893	0.734	0.448	0.24	-0.056	0.923
Share of household members age 0-13	0.797	0.160	0.327	0.176	-0.011	0.420
Household has electric light (=1)	0.348	0.448	0.313	0.192	0.074	0.348
Household had work animals (=1)	0.626	0.354	0.341	0.849	-0.055	0.107
Log distance to the municipality capital (km)	0.487	0.630	0.189	0.320	-0.117	0.569

Notes: *** p<0.01, ** p<0.05, * p<0.10. Compares baseline values from early treatment group (2000) with comparison group (2002). ATT biased adjusted estimator (Abadie and Imbens 2011) using five nearest neighbors. Standard errors on the differences are clustered at the locality level.

TABLE H3: 2010 ABSOLUTE IMPACT —ALTERNATIVE SPECIFICATION

	Education		Learning Family Z-Score	Economic Participation Family Z-Score	Earnings Family Z-Score		Socio-emotional Family Z-Score
	Grades Attained	Family Z-Score			Rank	Absolute (5% Trim)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: NN5 — Min-Max Trim							
ATT	1.375***	0.379***	0.386***	0.185	0.066	0.106	0.103
	(0.347)	(0.088)	(0.123)	(0.137)	(0.096)	(0.106)	(0.082)
N	690	690	616	690	690	687	613
Panel B: NN5 — 1 Percent Trim							
ATT	1.405***	0.387***	0.387***	0.184	0.070	0.108	0.096
	(0.349)	(0.089)	(0.124)	(0.133)	(0.094)	(0.102)	(0.083)
N	665	665	594	665	665	662	591
Panel C: NN5 —2 Percent Trim							
ATT	1.463***	0.398***	0.399***	0.209	0.070	0.108	0.095
	(0.334)	(0.087)	(0.124)	(0.133)	(0.091)	(0.100)	(0.084)
N	652	652	581	652	652	649	578
Panel D: NN2 — Min-Max Trim							
ATT	1.533***	0.425***	0.461***	0.255**	0.104	0.127	0.127*
	(0.303)	(0.072)	(0.095)	(0.120)	(0.088)	(0.108)	(0.067)
N	690	690	616	690	690	687	613
Panel E: Kernel Matching —Min-Max Trim							
ATT	1.447***	0.397***	0.414***	0.191	0.069	0.105	0.121**
	(0.434)	(0.100)	(0.116)	(0.154)	(0.115)	(0.110)	(0.059)
N	690	690	616	690	690	687	613
Panel F: Local Linear Regression Matching — Min-Max Trim							
ATT	1.280***	0.374***	0.377***	0.182	0.065	0.110	0.085
	(0.441)	(0.113)	(0.101)	(0.155)	(0.112)	(0.110)	(0.065)
N	690	690	616	690	690	687	613

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and are given in parentheses. Absolute effects compare early treatment to comparison group in 2010. Mean of grades attained in the comparison group is 5.3 in 2010 for panel A. Z-scores are calculated as the mean and divided by the standard deviation of the late treatment group.

APPENDIX I: NON-EXPERIMENTAL ABSOLUTE PROGRAM EFFECTS —DOUBLE DIFFERENCE

At some point during the period 2000-05, the three-year CCT operated in *all* rural areas of the six municipalities where the evaluation took place. This included the 42 rural localities randomized into early and late treatment as well as the 17 rural localities not initially targeted because they were less poor in 1995 according to the marginality index used for geographic targeting. In the additional 17 localities, the program began in 2001 and was offered to 80 percent of the population based on a household-level proxy means targeting model. Consequently, by 2005 the three-year program had been implemented (at different times and to modestly different degrees) in all 59 rural localities in the six municipalities, covering over 90 percent of the rural population in them. (See appendix C for further details.) Given this high coverage, it is possible to use national census data to provide evidence of absolute program effects, albeit for a somewhat limited set of educational and demographic outcomes available in the census.

Specifically, we use the 1995 and 2005 censuses and a non-experimental double difference approach to estimate absolute program effects five years after the CCT began in the early treatment areas. Together, the two censuses provide repeated cross sections at the individual level, in 1995 before the start of the program and in 2005, the year the program ended.[55] The censuses provide information on current municipality of residence (and whether rural or urban), as well as municipality at birth and municipality of residence five years prior to the census administration date.[56] We assign all individuals to the municipality where they lived 5 years prior to each census (about 3 percent of the 9-12 year olds moved in that period), and assume the type of prior residence (rural or urban) is the same as the current residence. We calculate double difference impacts using relevant age cohorts (calculating ages on November 1 as done for the main analyses, in 1990 and in 2000) and comparing outcomes in rural areas of the six program municipalities to outcomes in rural areas of the six neighboring municipalities where the non-experimental comparison group was selected in 2002. The 9–12 age cohort in 2000 is the same cohort examined in the main experimental analyses. In addition, we provide estimates for the 7-8 age cohort and 13-year olds, covering the entire age range of children potentially eligible for the education component of the CCT. We estimate

$$Y_{imt} = \delta_0 + \delta_1 T_{m,t-5} + \delta_2 C_t + \delta_3\, T_{m,t-5} * C_t + \varepsilon_{imt} \quad (1)$$

where Y_{imt} is the educational outcome for child i in municipality m measured in census year t, $T_{m,t-5}$ is an indicator for whether the child resided in a treatment municipality five years prior to the census year, and C_t an indicator for the 2005 census. δ_3 yields the double difference estimate five years after the program began on Y, which includes grades attained, enrollment and literacy.[57] Standard errors are robust to heteroskedasticity.

[55] It is not possible to link individuals across the two census rounds.

[56] While the data also include more detailed location information, changes in the definition and boundaries of census areas between 1995 and 2005 make it impossible to match them across time. Municipality boundaries, however, remained identical.

[57] δ_3 is the average impact for the three different groups—early treatment, late treatment and, the other 17 localities—each of which by 2005 had received the three-year program at different times. It is not possible to isolate the three distinct treatment group areas within the census data (see previous footnote).

Because the CCT did not operate in urban areas, all estimates limit the sample to individuals living in rural areas. The main double difference estimation equation takes a first difference between outcomes measured in 2005 for those living in program and non-program comparison municipalities in 2000, and a second difference between outcomes measured in 1995 for those living in program and non-program municipalities in 1990, as indicated by municipality of residence five years prior. These main results are presented in Table 11 in the text, and reproduced below.

In Table I1 we present the double difference impacts after five years for the three age cohorts (7-8, 9-12, and 13). Effects for the younger cohort are similar in magnitude and significance to the 9-12 year olds, consistent with the program having had a positive absolute effect on these boys as well. Evidence of effects for the 13-year olds is only a little weaker with highest grade attained, having completed fourth grade and enrollment all significant.

We also explore the sensitivity of the main double difference findings to different sets of comparison municipalities and definitions for treatment status (Table I2). First, we expand the comparison municipalities to include rural areas in all non-program municipalities in the central regions of Nicaragua where the program was located (panel B). Second, we examine whether results differ when we instead use current residence (panel C) or, separately, municipality of birth (panel D) to determine program eligibility. While there are some differences in point estimates (particularly for enrollment which ranges from 0.027 to 0.087), the different approaches all suggest significant absolute impacts of the program on educational outcomes after five years.

Last, we provide evidence in support of the identifying assumption by estimating program "effects" on outcomes for a different cohort unlikely to have been affected by the intervention—household heads in households with a child in the 9–12 cohort (using one observation per household).[58] The same empirical specification suggests there are no effects on their educational and demographic characteristics (Table I3). Moreover, point estimates are all close to zero, providing support for common trends.

[58] We analyze common trends using household heads rather than an older age cohort of children because older children may still have been influenced by the program, including through migration patterns that are difficult to disentangle using the national census data.

TABLE I1: 2005 ABSOLUTE IMPACTS ON EDUCATION AND CIVIL STATUS, BY AGE COHORT

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1	Ever Married =1
	(1)	(2)	(3)	(4)	(5)
δ3 for 7-8 age cohort	0.479***	0.104***	0.032*	0.084***	-0.001
	(0.076)	(0.018)	(0.018)	(0.017)	(0.001)
δ3 for 9-12 age cohort	0.597***	0.124***	0.037***	0.091***	-0.003
(*as in Table 11*)	(0.078)	(0.014)	(0.014)	(0.013)	(0.004)
δ3 for 13 age cohort	0.494**	0.078**	0.046*	0.036	-0.024
	(0.198)	(0.031)	(0.025)	(0.029)	(0.019)
N for 7-8 age cohort	11,056	11,056	11,068	11,045	11,067
N for 9-12 age cohort	18,399	18,399	18,421	18,403	18,421
N for 13 age cohort	4,144	4,144	4,148	4,145	4,148
N Total	33,599	33,599	33,637	33,593	33,636
Mean comparison group 7-8	2.965	0.416	0.709	0.765	0.613
Mean comparison group 9-12	3.922	0.559	0.456	0.779	0.282
Mean comparison group 13	4.321	0.580	0.271	0.746	0.223
P-value (7-8 vs 9-12)	0.282	0.376	0.833	0.523	0.053
P-value (9-12 vs 13)	0.623	0.179	0.764	0.627	0.191

Notes: *** p<0.01, ** p<0.05, * p<0.10. 2005 absolute effects use national census data to compare rural areas of program municipalities to rural areas of the six comparison group municipalities. Birth information unavailable for 7-8 year old cohort (too young at time of census) and unreported (missing) for 7-17 percent of observations in other age cohorts. Heteroskedasticity-robust standard errors are given in parentheses. All variables measured using the 1995 and 2005 population censuses and include boys at the indicated ages in November of 1990 and 2000, respectively.

TABLE I2: 2005 ABSOLUTE IMPACTS ON EDUCATION AND HOUSEHOLD HEAD CHARACTERISTICS

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1
	(1)	(2)	(3)	(4)
Panel A: Treatment Municipality (5 Years Prior) vs 6 Municipality Comparison Group				
Treatment municipality * 2005				
(δ_3)	0.597***	0.124***	0.037***	0.091***
	(0.078)	(0.014)	(0.014)	(0.013)
N	18,399	18,399	18,421	18,403
Mean comparison group 2005	3.922	0.559	0.456	0.779
Panel B: Treatment Municipality (5 Years Prior) vs Central Regions Comparison Group				
Treatment municipality * 2005				
(δ_3)	0.537***	0.101***	0.087***	0.085***
	(0.055)	(0.010)	(0.010)	(0.009)
N	85,903	85,903	86,034	85,884
Mean comparison group 2005	3.772	0.537	0.398	0.755
Panel C: Treatment Municipality (Current) vs 6 Municipality Comparison Group				
Treatment municipality * 2005				
(δ_3)	0.646***	0.129***	0.038***	0.091***
	(0.078)	(0.014)	(0.014)	(0.013)
N	18,324	18,324	18,348	18,332
Mean comparison group 2005	3.896	0.556	0.453	0.779
Panel D: Treatment Municipality (of Birth) vs 6 Municipality Comparison Group				
Treatment municipality * 2005				
(δ_3)	0.552***	0.113***	0.027*	0.084***
	(0.079)	(0.014)	(0.014)	(0.014)
N	18,206	18,206	18,232	18,210
Mean comparison group 2005	3.860	0.547	0.445	0.771

Notes: *** p<0.01, ** p<0.05, * p<0.10. 2005 absolute effects use census data to compare rural areas of program municipalities to rural areas of other municipalities. Heteroskedasticity-robust standard errors are given in parentheses. All variables measured using the 1995 and 2005 population censuses and include boys at the indicated ages in November of 1990 and 2000, respectively.

TABLE I3: 2005 ABSOLUTE IMPACTS ON HOUSEHOLD HEAD CHARACTERISTICS

	Grades Attained	Completed Grade 4 =1	Read and Write =1	Female Head =1	Age of Head in Years
	(1)	(2)	(3)	(4)	(5)
Treatment municipality * 2005 (δ_3)	-0.020	0.014	0.005	0.007	0.342
	(0.079)	(0.013)	(0.016)	(0.013)	(0.417)
N	15,192	15,192	15,286	15,292	15,292
Mean comparison group 2005	1.673	0.226	0.507	0.221	46.901

Notes: *** p<0.01, ** p<0.05, * p<0.10. 2005 absolute effects use census data to compare rural areas of program municipalities to rural areas of other municipalities. The mean of the comparison group is for the six comparison group municipalities. Sample includes household heads in households with a child in the 9-12 age cohort. Heteroskedasticity-robust standard errors are given in parentheses. All variables measured using the 1995 and 2005 population censuses and include heads from households with a boy or girl age 9-12 in November of 1990 and 2000, respectively.

APPENDIX J: COST-BENEFIT CALCULATIONS

As described in the text, a full cost-benefit analysis would require monetary assessment of all of the resource costs as well as all of the potential (positive or negative) effects of the CCT. Assessment is further complicated because we can only estimate differential effects with the experiment and do not have experimental evidence on absolute effects, although the evidence strongly suggests for outcomes considered in this paper they are at least non-negative. In this appendix we outline the approach taken for approximating the net present value (NPV).

Following the approach outlined in Dhaliwal et al. (2012), all values are first translated into U.S. dollars using market exchange rates after which they are deflated to 2000, the base year, using U.S. CPI. Costs included are direct program costs for running the program (administrative/management, targeting, monitoring and conditionality) but as recommended for cost-benefit analysis do not include the transfers or evaluation costs (reported in Caldés and Maluccio 2005). We allocate 50 percent of the average per household program cost as an estimate of costs related to the education components of the program, in nominal terms $60 a year for three years. We assume that the other 50 percent were costs related to the nutrition and health components of the program. Benefits are calculated as the estimated average increase monthly earnings ($10) starting in 2009, multiplied by average number of months worked (3.5) and by the average number of men in the cohort in each household (1.37). All these dollar figures are first deflated to year 2000 constant U.S. dollars and then discounted at 5 or 10 percent discount rates. Using a 10 percent discount rate, the NPV turns positive in 2027, approximately 20 years after the program ended; at 5 percent, the NPV turns positive in 2016.

In the above, we do not directly account for other important potential costs, however, related to the potential deadweight losses associated with taxation necessary to make the transfers and the possibility that the CCT crowds out other governmental expenditures with higher returns (Harberger 1997). Assuming households received full education transfers for three years, and that the cost of raising those funds was 5 percent (at a continued discount rate of 10 percent), NPV would take an additional eight years to turn positive, in 2035.

While it is impossible to be certain there are no negative impacts of the program we did not capture that should be treated as additional costs, at each discount rate arguably the estimates are a lower bound on NPV for several reasons. First, we attribute *all* of the costs of the education component to a subset of the eligible beneficiaries (but they were relevant also for eligible boys 7-8 and 13, as well as for girls 7-13). Second, we consider benefits only of one type, increased off-farm earnings. For example, this only includes benefits from the schooling and learning gains that operate directly through the labor market. And while off-farm earnings gains in theory might have been offset by reduced earnings on the family farm, the estimates show the CCT did not decrease work on the family farm on the extensive margin so that a reduction in earnings is unlikely. Third, we exclude any potential gains for the younger and older age cohorts of boys, as well as for girls. Analyses reported in this paper for boys of different ages and in Barham, Macours, and Maluccio (2018) for girls all point to positive, or at least non-negative, effects on labor market outcomes for these other groups as well. Fourth, we only consider benefits starting in 2009. Last, we ignore program costs associated with boys in the late treatment group further offsetting the costs of program in early treatment in the differential framework.

Overall then, and while admittedly uncertain, under relatively conservative assumptions regarding cost and benefit flows for a subset of the beneficiaries, we conclude the program could achieve positive NPV within a few decades.

Additional Appendix References

Abadie, A., and G. Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica* 74(1): 235–267.

Adato, M. and T. Roopnaraine. 2004. "A social analysis of the Red de Protección Social. Report submitted to the *Red de Protección Social*." International Food Policy Research Institute. Washington, DC.

Adhvaryu, A., Molina,T., Nyshadham, A., and Tamayo, J. 2018. "Helping Children Catch Up: Early Life Shocks and the *PROGRESA* Experiment." Mimeo, University of Michigan.

Angelucci, M. and G. De Giorgi. 2009. "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?" *American Economic Review* 99(1): 486–508.

Angelucci, M., G. De Giorgi, and I. Rasul. 2017. "Consumption and Investment in Resource Pooling Family Networks." *Economic Journal*, forthcoming

Angelucci, M., G. De Giorgi, M.A. Rangel, and I. Rasul. 2010. "Family Networks and School Enrolment: Evidence from a Randomized Social Experiment." *Journal of Public Economics* 94: 197–221.

Arcia, G. 1999. "*Proyecto de Red de Protección Social: Focalización de la fase piloto*" Report to the Inter-American Development Bank, Washington DC.

Attanasio, O., S. Cattan, E. Fitzsimon, C. Meghir, and M. Codina. 2015. "Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia." *IFS Working Paper* 15/06.

Blattman, C., N. Fiala, S. Martinez. 2014. "Generating Skilled Self-employment in Developing Countries: Experimental Evidence from Uganda." *Quarterly Journal of Economics*: 697-752.

Caldés, N., and J.A. Maluccio. 2005. "The Cost of Conditional Cash Transfers." *Journal of International Development* 17(2): 151–168.

Cunha, F., J.J. Heckman, and S.M. Schennach. 2010. "Estimating the Technology of Cognitive and Non-Cognitive Skill Formation." *Econometrica* 78(3): 883–931.

Dehejia, R., and S. Wahba. 1999. "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs". *Journal of the American Statistical Association* 94 (448): 1053–1062.

Dhaliwal, I., E. Duflo, R. Glennerster, and C. Tulloch. 2013. "Comparing Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education." in P. Glewwe, Ed. *Education Policy in Developing Countries*, Oxford UK: Oxford University Press.

Doyle, O., C. Harmon, J.J. Heckman, C. Logue, and S.H. Moon. 2017. "Early skill formation and the efficiency of parental investment: A randomized controlled trial of home visiting." *Labour Economics*, *Labour Economics*, 45:40-58.